# Royal Australian Navy Meteorology and Oceanography Operations Research Report

Susan S. Kirschenbaum
Combat Systems Department

NAVSEA
NEWPORT
Undersea Warfare Center Division
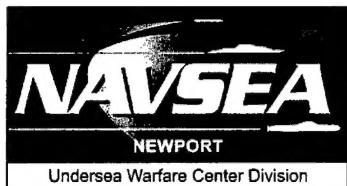
Naval Undersea Warfare Center Division
Newport, Rhode Island

20020508 114

PREFACE

Reviewed and Approved: 25 January 2002

Philip A. LaBrecque
Head, Combat Systems Department

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE 25 January 2002	3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Royal Australian Navy Meteorology and Oceanography Operations Research Report	

6. AUTHOR(S)

Susan S. Kirschenbaum

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Naval Undersea Warfare Center Division 1176 Howell Street Newport, RI 02841-1708	TR 11,346

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
Office of Naval Research Ballston Centre Tower One 800 North Quincy Street Arlington, VA 22217-5660	

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Approved for public release; distribution is unlimited.	

13. ABSTRACT (Maximum 200 words)

This report documents the results of a study of Royal Australian Navy weather forecasters. The results replicate and extend the accomplishments of the ONR-funded Meteorology and Oceanography (METOC) Human Systems Interaction Improvement (HSII) Project. The objective was to map information usage of U. S. Navy (USN) METOC forecasters to visualization tools, and compare that mapping to mappings employed by another group of forecasters, located in a very different geographic region, employing different tools, and with different patterns of staffing and training. Data collected at two Royal Australian Navy (RAN) facilities replicated the principal findings of HSII in terms of workflow and extraction from complex visualizations. They extended the findings by documenting how forecasters assess the associated uncertainties in models and data.

14. SUBJECT TERMS		15. NUMBER OF PAGES 48
Weather Forecasting	Process Tracing	
Protocol Analysis	Assessment of Uncertainty	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT SAR

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

Section   **Page**

LIST OF ILLUSTRATIONS

Figure   **Page**

LIST OF TABLES

ROYAL AUSTRALIAN NAVY METEOROLOGY AND OCEANOGRAPHY OPERATIONS RESEARCH REPORT

1. INTRODUCTION

This report documents the results of a set of experiments conducted in cooperation with the United States Navy (USN) and the Royal Australian Navy (RAN). The objectives were threefold: First, to map the information usage of the Meteorology and Oceanography (METOC) decision maker to information visualization tools. Second, to compare the mappings of USN and RAN forecasters in order to distinguish between the effects that are dictated by the tools and training of these specialists and those that are due to basic human cognition. Third, this research extends the USN results to additional exploratory findings. The results reported here replicate and extend the work performed under the METOC Human System Interaction Improvement (HSII) Project sponsored by the Office of Naval Research. Research team members include James Ballas, Gregory Trafton, Susan Kirschenbaum, Robert Miyamoto, Sandra Marshall, and Nicholas Gizzi Jr. In the HSII experiments, data were collected in semi-structured experimental conditions at the Naval Pacific METOC Center (NPMOC) at the Naval Air Station (NAS), North Island, San Diego, and aboard USS *Carl Vinson*; the data were presented at the Proceedings of the Human Systems Integration Symposium in Garden City, Virginia.[1]

This work is a member of a class of cognitive research that seeks to establish how highly trained and experienced personnel interact with assorted specialized tools and visualizations to perform complex tasks. The first step in research of this type usually entails observational studies of the domain personnel working in both realistic and simulated settings. The studies are videotaped and then analyzed for process, tool, and visualization use. Finally, models are created of the process, which serve as hypotheses for subsequent rounds of experimentation. These procedures have been pioneered by numerous researchers[2,3,4] in domains as diverse as puzzle solving (Tower of Hanoi[5]), economics problem solving,[6] computer programming,[7] and that of the long-distance telephone operator.[8]

This report also addresses how people understand and use visual representations of all kinds to make decisions. Notably, experts incorporate both physical objects and elements of their expertise (e.g., force vectors[9] or other symbols[6]) and also employ specialized tools[10] in understanding how people use visual representations in the decision-making process.

1.1 SPECIFIC PROBLEMS

1.1.1 Replication

For the past several years, a team of researchers has been investigating how experienced METOC forecasters use available information and tools to perform their tasks in a timely and

accurate manner. The forecast process and the impact of tools, especially visualizations, are the foci of this research. To date, a hypothesized workflow description, an inventory of tools, and a high-level description of the process have been developed.

Major concerns of this work lie in understanding how METOC forecasters develop mental representations of the highly complex and dynamic forces at work in the atmosphere, and how they integrate static satellite pictures, predictive models, and their own knowledge to create accurate predictions. Major complications include making modifications to compensate for mis-matches between model predictions and measured data, accounting for sparse and uncertain data, and interacting with a variety of visualizations, including satellite pictures, interactive model graphics, and alphanumeric data readouts. The first experiment is detailed in Trafton et al.[11] The most striking finding is that forecasters make extensive use of qualitative visualizations to build a mental model of the weather systems, including predictions for the target period. During this period, significantly more qualitative information is extracted than quantitative; however, when preparing a forecast brief or report, forecasters use far more quantitative language (e.g., changing words such as hot to 36°C).

The generalizability of these findings and the impact of different meteorological and cultural environments, training, and military toolsets will extend this research and support a more robust effort to validate the mental model.

1.1.2 Extension: Assessing Uncertainty

The extension reported in this report addresses the assessment of uncertainty in METOC data, both in observations and models. Neither the USN nor the RAN forecasters explicitly mentioned uncertainty or probability; however, uncertainty is inherent in weather data, and forecasters must make judgments that account for and accommodate that uncertainty. This assessment and accommodation is all the more difficult because the level (or even presence) of uncertainty is not displayed in any of their tools.

Decision making (forecasters are decision makers) is said to occur under uncertainty in any situation in which the decision maker does have perfect knowledge about the true state of the world. Thus, weather forecasting is an example of decision making under uncertainty. There are two classes of decision making under uncertainty. In the first, decision makers have access to some representation of the amount of uncertainty in their information. This kind of situation has received considerable attention in the scientific literature[12, 13, 14] for quite some time. The second class of decision making under uncertainty occurs when there is no information given concerning the magnitude of the uncertainty, or the information is not available in the source data. This situation has received virtually no attention in the scientific literature and is the case with weather forecasting.

All decision makers act on second-hand information. The "truth" is transformed in many possible ways (e.g., direct transformation, modeling, a combination of the two, or multiple representations). Weather forecasters view reports of observations, not the observations themselves. They see satellite images and loops uploaded from the Internet. They view

2

predictions made by complex models that contain varying underlying assumptions. (The models, of course, are also based on possibly unreliable or sparsely sampled observations.) The associated uncertainties, unreliabilities, data sparsity, and assumptions are not explicitly provided to the forecaster. She or he must infer the magnitude and direction of the uncertainty from such sparse data as observing that the values are not stable across time; that the predictions a given model made for *now* is inaccurate (too fast, too far north, too high, etc.); or that different instances of the "same" data (e.g., different weather models) make different predictions.

Not only are the probabilistic relationships between the events in the world and the stimulus information often not available, even the visualizations do not just appear on the display surface. The forecaster must search for them or select information to view them from a much larger set of options. As he or she must select what information to view, tracing the information gathering process provides an insight into the decision process, including strategies for both assessing and accommodating the uncertainty. The decision maker's knowledge about the domain guides information-selection strategies. Thus, expertise plays a large role in information gathering, especially under time constraints. Experts tend to view less total information than novices;[15, 17] however, experts do scan all categories of information.[16, 17] Novices, on the other hand, can develop tunnel vision when there is critical action to be taken,[16] failing to scan possibly relevant items.[17] Time limits may not permit the investigation of all the information, but when time limits permit, the prudent decision maker delays taking action until all the stimuli have been sampled and until the information is available to reduce the uncertainty.

1.2 ORGANIZATION OF THIS REPORT

Section 2 is an introduction to the RAN METOC facilities at the three primary sites where the research was conducted. It documents the Organization, Tools, Procedures, and Training that are common for all facilities. Section 3 provides information on the tools and training of the USN METOC community and is provided for comparison purposes. Section 4 describes the data collection procedures, including special procedures. Section 5 provides the data analysis procedures, and section 6 contains the results of the research. Section 7 provides the conclusions and recommendations. Section 8 lists references cited in the text. A chronological trip report and other supporting documentation are given in the appendix.

2. METOC ORGANIZATION, TOOLS, FACILITIES, AND TRAINING

2.1 ORGANIZATION AND RESPONSIBILITIES

The METOC organization of the RAN is quite small, given the size of the region that the ships patrol and the Australian area of influence. RAN METOC responsibilities cover the naval ports in Australia and various regional waters, ranging from Antarctica in the south to trouble spots in Oceania and southeast Asia in the north. RAN METOC operations also support exercises, including those with international partners. There are naval bases in tropical Darwin

and subtropical Sydney, in addition to the temperate latitudes. To the east, RAN METOC works closely with its nearest neighbor, New Zealand, to patrol the Tasman Sea. To the west, a vast expanse of ocean lies between Australia and Africa. The western-most naval base is the HMAS Stirling in Garden Island, Rockingham, in Western Australia, which is located near Perth. With the wide geographic variation and the range of climates and ocean conditions, the responsibility of RAN METOC is extensive.

While there are no active aircraft carriers in the RAN fleet, there are planes and helicopters that perform missions ranging from rescue to training to searching for submarines. These aircraft are home based at the HMAS Albatross, Nowra, New South Wales. The HMAS Stirling is the principal submarine base. There are other bases that serve as home ports to a variety of naval vessels. Australia also frequently hosts visiting ships from allied nations, including U.S. aircraft carriers.

The Directorate of Oceanography and Meteorology is headed by the Director of Meteorolgy (DOM), a professional naval oceanographer/meteorologist, who manages the Directorate and serves as the RAN METOC expert for numerous national and international organizations. The DOM also maintains strong ties to the civilian Bureau of Meteorology (BoM). There are two major divisions under the director, the Australian Oceanographic Data Centre (AODC) and the Operational METOC Centre (OMC), both located in Garden Island, Sydney. The AODC is responsible for the development and management of data acquisition, manipulation, and visualization capabilities, including the Defence Oceanographic Data Centre, which is headed by a senior defense civilian. The two primary activities are research and development of new tools, and database management and validation. The latter also includes significant research and development.

The OMC, headed by a naval METOC officer, coordinates the two forecasting centers (although each center has an officer in charge (OIC)). From the OMC, METOC professionals are sent to operational assignments, as needed. Additionally, much research and development is also coordinated through this office, and METOC training is conducted. One of these METOC centers, the Fleet Weather and Oceanography Centre (FWOC), is located in the Maritime Headquarters building in Garden Island, Sydney. It supports at-sea and in-port planning and fleet operations. The other METOC center, the Naval Air Station Weather and Oceanography Centre (NWOC), is located at the HMAS Albatross, the airbase in Nowra, New South Wales. This center supports air operations. In both cases, their location puts the METOC centers close to their primary customers: the Maritime Command and planners for FWOC, and the flight commanders and the flight tower for NWOC.

2.2 TOOLS

The majority of the tools are web based. Therefore, they can be displayed on any networked computer (with the appropriate clearance if the tool displays classified data). The Australian BoM works closely with the RAN to develop and support the required databases and tools. Additional tools are under development, but are not in common use.

The following are the major tools used by RAN forecasters:

- MCIDAS is the principal weather forecaster's tool. It can display a multitude of geographically referenced information and is most often used with a satellite picture or satellite loop displayed. Exact location and data values can be read from the location of the cursor. A tool named Kenny provides additional visualization options for the basic MCIDAS data set.

- TESS-II is used for forecasting and long-range planning for ocean acoustics. It also displays conditions graphically after the user selects the location, date/time, sonar system(s), targets, etc. This tool is easier to use than the products available in the USN's Submarine Fleet Mission Program Library (SFMPL). It selects the optimal algorithm, based on anticipated range, time to calculate, etc.

- TURBOWIN is used to aid the non-METOC weather observer who is usually on board the ship. TURBOWIN's use improves observatory accuracy, it is faster than manual logging, and it automatically generates the coded weather message in the format used by the larger databases and models.

- OCEANS is based on a U.S. product, ARC Info, that was originally developed for the Army. It is used to display all geographical information, including observations, model output, vector data, etc. Currently, the available databases are primarily oceanographic, not weather based. The information is tiled into as many windows as necessary to display the selected data. Individual numeric data can also be viewed, and may be modified and redisplayed.

- Reporting tools include charts, overhead projector slides, and computerized programs. New tools, some locally developed, aid the forecaster in recording and disseminating the forecast. When the forecast is recorded and updated on a networked computer, it is accessible to all customers at their convenience and at a distance. Forecasts are briefed by the forecasters to local users and sent by fax and computer (E-mail and a shared reporting system, such as the terminal aerodrome forecast (TAF)) to distant users.

Displays are also available for the local radar, the TAF, and local observations. Charts and predictions from the BoM and other sources are printed for reference. In addition, each forecaster develops one or more daily analysis charts for the entire continent and vicinity. These charts serve as a record of the weather conditions and are the starting point for forecasts and briefs. They are developed from on-site observations and validated by satellite images. The process of developing these analysis charts will be described in detail in section 5 of this report.

Each observation must be validated for reasonableness. Currently, this is done by unaided inspection. A new tool under development must validate observations prior to their acceptance into the official database. The new tool will help distinguish anomalies from seemingly valid observations. For example, there was a recent XBT observation where the recorder apparently wrote a "3" rather than a "5" in noting the ship's position. This put the ship over land rather than at sea. There are many other examples where problems have occurred. The graphical display makes the individual's job of performing the validation much easier and faster. Outliers and anomalies stand out when shown grahically, side by side with other data points.

Tools such as OCEANS and MCIDAS depend on models for their data. RAN forecasters have access to numerous models. The ones used most often are produced in Australia, Europe, and the United States. The high-resolution models, produced in Australia, are limited in their projections into the future. Therefore, other models are used for long-range forecasting. The selection of models depends on current conditions. Analysis data track which model performs the best at predicting given particular conditions. All appear to have strengths and weaknesses.

2.3 FACILITIES

Figures 1 and 2 show plan and elevation views of the forecaster's workstation at FWOC. FWOC is a secure space and therefore has no windows. As shown in the figures, there are multiple dedicated computers. Although each is dedicated to only a few tools, they are largely interchangeable. Thus, if one becomes inoperable or is unavailable, another can be substituted. Because there are a number of computers with network connections, forecasters usually do not need to redirect their Internet browsers to access data.

In contrast to FWOC, NWOC (see figure 3 for the layout) is responsible for making regular weather observations and forecasting for locally based aircraft. These tasks are facilitated by the two walls of windows that face in the direction of much of the approaching weather and the runway. The runway itself is not visible from the windows due to other structures. It is, however, co-located with the airfield tower. The buildings where the squadrons are based and the hangers are located across the runway from the METOC office.

6
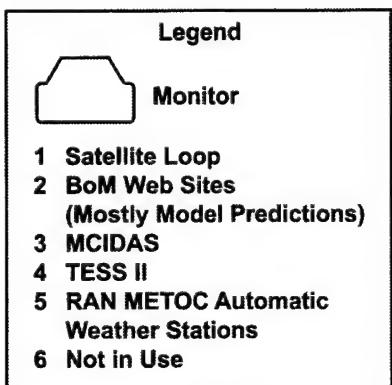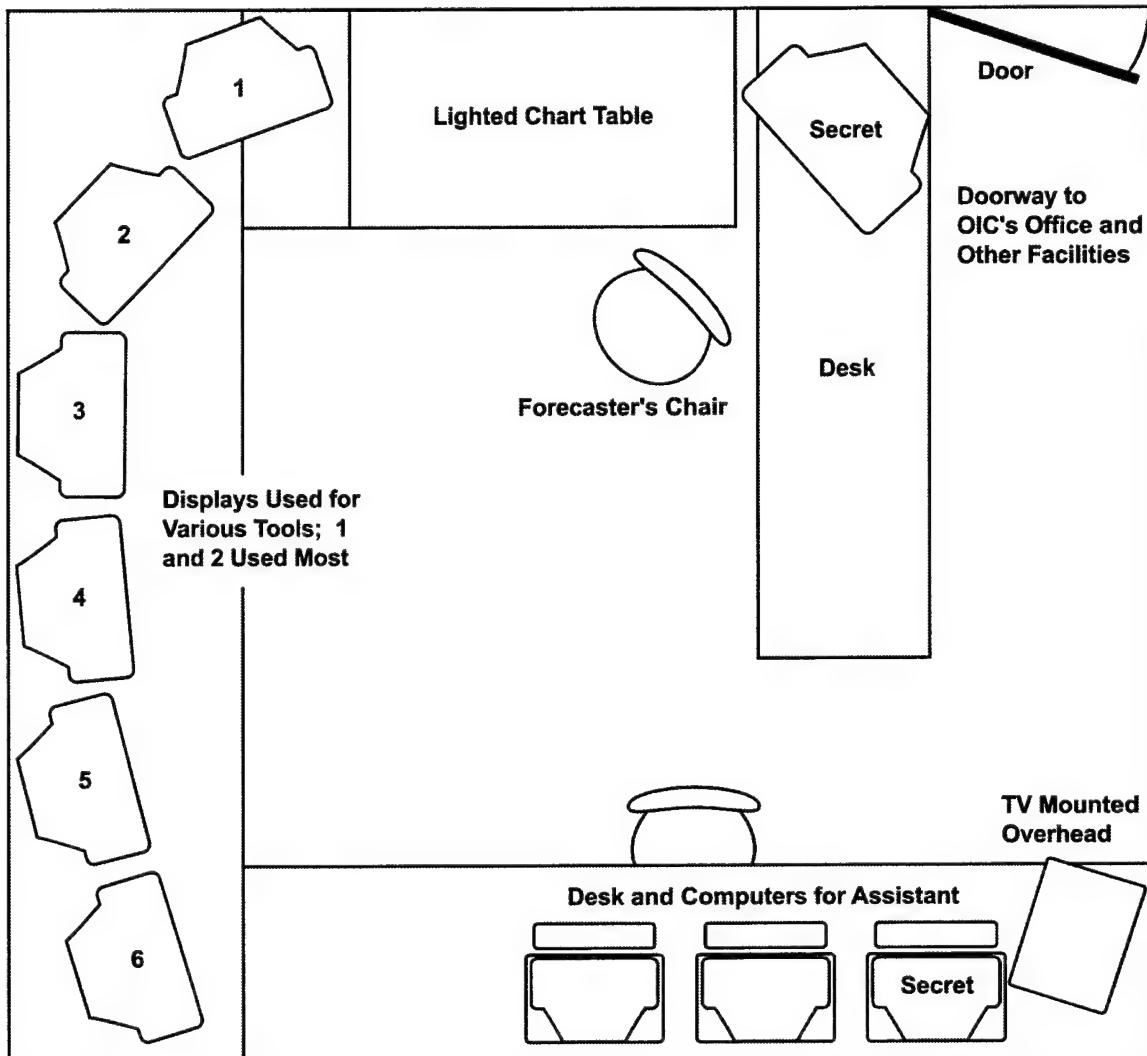
Figure 1. Plan View of Forecaster's Station at FWOC
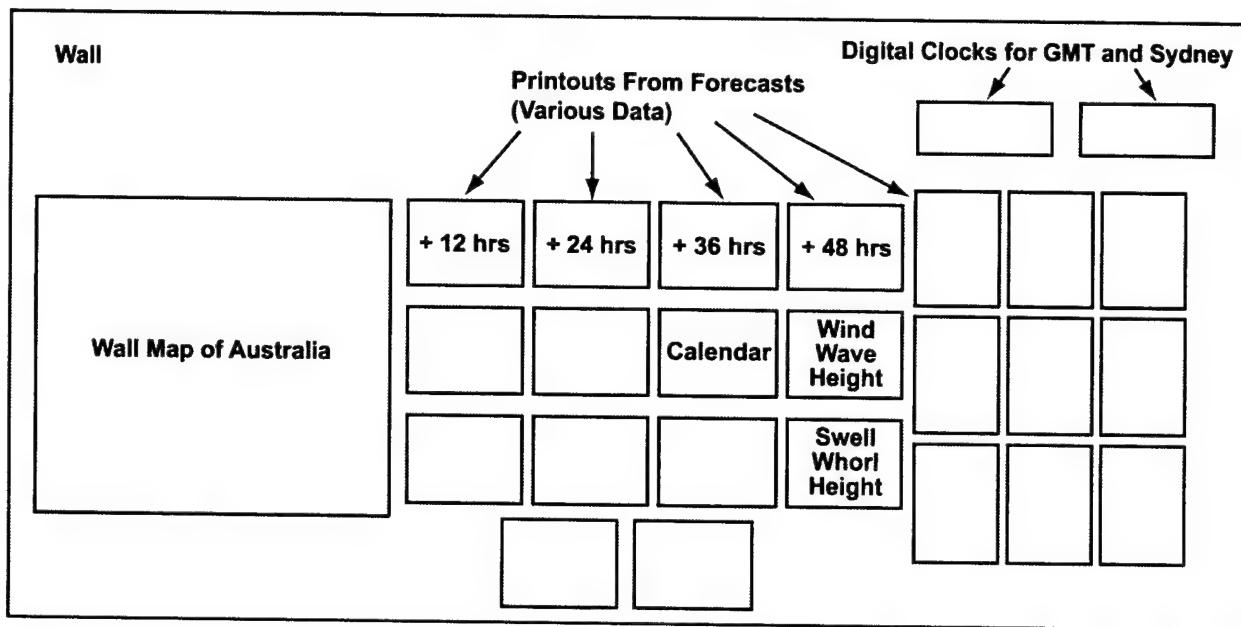
Figure 2. The Wall Above the Forecaster's Station at FWOC

Windows Run the Full Length of the Wall

Copy/Fax

Observation
Displays Used for
Recording and
Accessing Current
Conditions

Display Boards
for Current
Forecast and
Weather
Information

Plotting Table

2-Headed Computer

Printer

Window in Wall

OIC's Office

Monitors
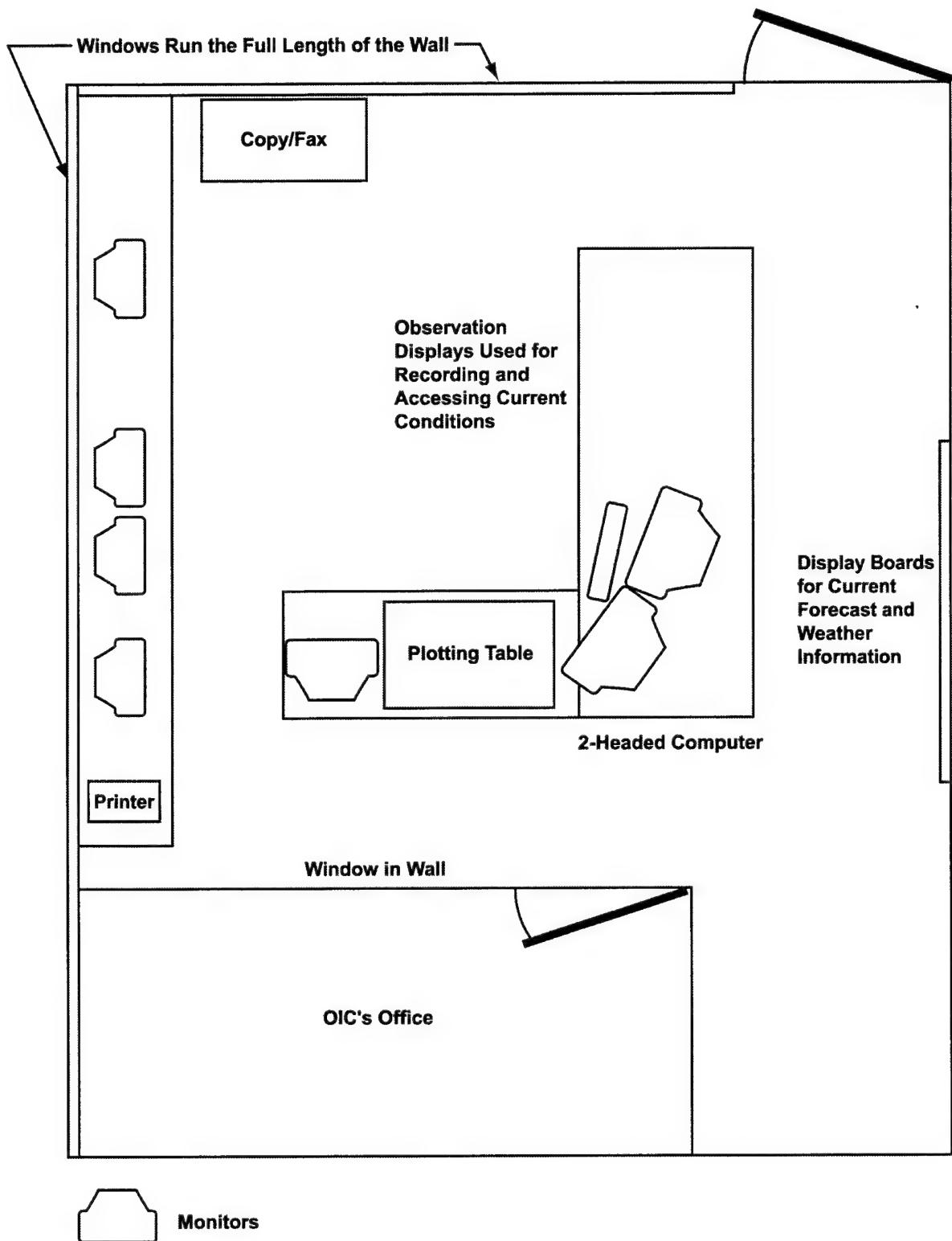
Figure 3. Plan View of NWOC Forecaster's Workstation and Facilities

2.4 TRAINING

Forecasters are all naval officers with a minimum education of a university degree at the Bachelor's level. Subsequently, they are trained by the Australian BoM in a 9-month meteorology course and then complete a 3-month naval course emphasizing tactical implications. Upon completion of these courses, they receive a diploma from the BoM. The academic cirricula are based heavily on mathematics and physics. Trainees also practice all aspects of weather forecasting, such as reading charts and satellite images and producing analysis charts. At the time of this study, the forecasters at FWOC ranged from newly trained to those with 2 years experience. Those at NWOC ranged in experience from about 1-month to several years. The OIC has significantly more experience. Thus, forecasters range from novices to journeymen, and the OIC is generally at the expert level.

2.5 STAFFING

Table 1 summarizes the staffing at both facilities during the observation period, and does not reflect authorized billets or a typical staff level. At FWOC, the OIC served in a supervisory and training/mentoring role for the forecasters. At NOWOC, the OIC also took the forecasting duty for one rotation.

Table 1. Staffing at FWOC and NWOC During Observation Period

Staff	FWOC	NWOC
IOC	1	1
Forecasters	3	2
Technicians	3	0

At FWOC, there was one forecaster and one technician on duty at all times. The OIC is the senior member of the METOC staff at each facility and is usually in the office during normal working hours. FWOC operates 24 hours a day, 7 days a week with a small staff of forecasters who prepare three briefs a week in addition to meeting all operational requirements. Briefs are prepared using PowerPoint software and are usually briefed from an electronic copy that is E-mailed to the briefing location. In one observed instance, a novice forecaster was presenting her brief to the Admiral. The PowerPoint graphics failed to work; however, she was well prepared and able to improvise. She was observed phantom-pointing and looking to various locations on the imagery map while briefing. This is an indication that she was using mental imagery and recalling the graphics.

NWOC is staffed during the work day from approximately 0500 to 1800. The hours cover scheduled flights but are adaptable and may run longer. NWOC is staffed during the weekend only if flights are scheduled. Staffing includes the OIC and one forecaster. At the time of these observations, NWOC was short staffed. There were only two forecasters, in addition to the OIC. There are billets for additional forecasters and technicians, but none were assigned at that time.

3. FOR COMPARISON PURPOSES: USN TOOLS AND TRAINING

The following general information is provided to enable the reader to understand the differences between the tool sets and training of the RAN and USN participants. Specific differences between the data collection processes will be detailed below, where appropriate.

3.1 USN TOOLS

While many tools are available, there are several that are used most frequently. These include an extensive set of websites, the Joint METOC Viewer (JMV), specific tactical decision aids (TDAs), and PowerPoint software. The websites provide observations and model data, including satellite pictures of various types (loops, infrared images, etc.), TAFs, and forecast models at various scales. Some products are locally centered, while other centers extend their products to the national and international range. Most have limited access.

Model data sets can be downloaded to the local computer and visualized using the JMV. The JMV adds flexibility and usability to modeled data sets by allowing the forecaster to select from a very large number of data sets for simultaneous viewing. For example, the user could view wind speed and direction at different altitudes over a specified geographic region, or could view wind speed/direction, temperature, and humidity at a single altitude. Up to five variable sets can be combined in a single visualization.

TDAs are used to assess the impact of current and predicted weather conditions on missions. Additionally, they may be used to evaluate the effectiveness of proposed observation and weapon technology and, indirectly, recommend tactics to mitigate the effects of weather. PowerPoint software is the most common reporting tool. Different commands (and forecasters) have templates for developing their briefs. These can either be accessed by users online or briefed by the forecaster.

3.2 USN METOC TRAINING

There are two levels of USN forecasters: enlisted and officers. The enlisted rate is Aerographer's Mate (AG). AGs must have a minimum of a high school diploma. They attend a series of schools that train them to be weather observers and forecasters, and are certified as a weather forecaster and oceanographic specialist by both the Department of Defense and the Department of Commerce. AGs develop the majority of weather forecasts and are supervised by METOC officers.

To be a METOC officer (1800 code) requires a physical science degree (preferably meteorology or oceanography) with a strong background in physics and calculus. The career path for an 1800 officer usually includes an initial tour as an unrestricted line officer (ship driver) before converting to the 1800 code. Officers then attend graduate school for a master's degree in meteorology and oceanography. Approximately 10% earn doctorates (Ph.Ds).

4. DATA COLLECTION PROCEDURES

There were no experimental manipulations conducted in the RAN studies. Forecasters simply performed their normal tasks. However, they were requested to talk aloud while working. This procedure is common in observational data collection and provides some access to the cognitive processes of the forecasters.[2] Where possible, sessions were videotaped. When security requirements precluded taping, the experimenter recorded visualizations and the associated verbalizations in MacShapa,[18] which is a computer tool that facilitates time-tagged observations either in real time or from videotapes. The online recording is not subject to revision or inter-rater reliability computation and therefore is less reliable. The tape-recorded sessions could be encoded by more than one individual, which allows inter-rater reliabilities to be calculated.

All forecasters read the experimental protocol and agreed to participate in the study before the videotape was started. The forecasters were then instructed to talk out loud while they worked. If they were silent for an extended time, they were reminded to keep talking. At FWOC, 15.9 hours of observations were recorded real time, without taping. At NOWC, all on-task activities were videotaped. This resulted in 16.5 hours of observations and videotapes. On-task activities included taking observations, drawing charts, making forecasts, briefing customers, answering questions from walk-in and telephone customers, and discussing the current and predicted conditions with colleagues and supervisors. No personal activities such as personal phone calls, eating, etc., were recorded, although task interruptions were noted. Thus, a personal phone call that interrupted the forecaster while drawing a chart was recorded as an interruption, although it was not videotaped. The video recorder was mounted on a tripod and positioned to record the work surfaces (chart table and/or computer screen) and not the forecaster. However, as the forecaster moved around the center, the video camera followed. In order to gain a better understanding of the routine, supporting facilities, and responsibilities, each forecaster was interviewed regarding the tools and tasks.

5. DATA ANALYSIS PROCEDURES

5.1 GENERAL DESCRIPTION

Data were analyzed at three levels. At the least detailed level, data were analyzed to trace the workflow. This level of analysis globally categorizes the forecasters' tasks and extracts the pattern of the flow of work from one task to the next, across forecasters and settings. Workflow analysis provides a framework for the more detailed analyses.

At the more detailed level, a cognitive task analysis (CTA)[18] links work stages to the supporting cognitive tasks, such as extracting data from sources. Comparing CTA results for USN and RAN forecasters provides the data framework for evaluating the impact of tools and training on the cognitive process. Where there were similar results despite tool and training differences, support was found for attributing the results to similarities in the basic human cognitive processes.

Finally, a tool/task analysis was used to evaluate how well the tools supported the tasks. The tool/task metric (figure 4) is the proportion of time in the tool-only space, i.e., manipulating a tool without advancing the main task versus the total time (Tool $\cup$ Task). For example, if, while writing a report, the author changes fonts to emphasize a point, this advances the primary task. If, however, the author needs to customize the toolbar so that the font tool is available, that merely changes the word processing tool without advancing the writing task.
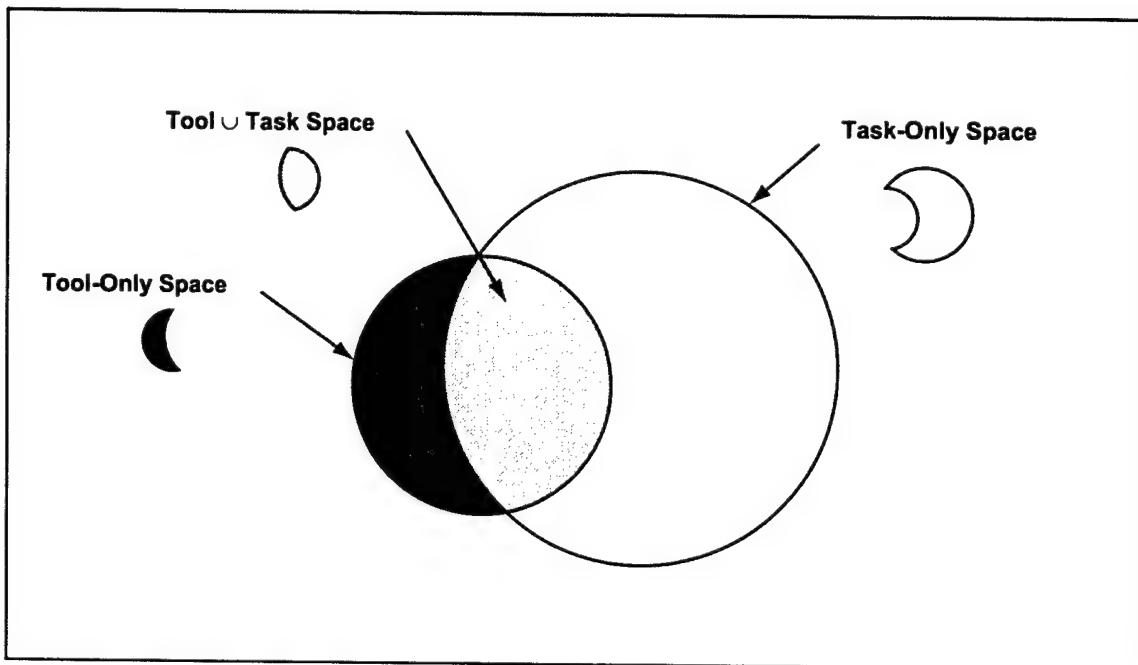
Figure 4. Diagram of Tool $\cup$ Task Space and Component Parts

5.2 VIDEOTAPED PROTOCOL ANANYSIS

Analysis of videotaped verbal protocols has been discussed extensively by Ericsson and Simon[2] (see also Gray and Kirschenbaum[20]). It is a time consuming and tedious process, but it provides insight into both cognitive processes and human-computer interactions. The outcome is the critical CTA that documents the interaction between the environment (input and output) and human cognition.

14

Analysis of the 15.9 hours of observations at FWOC and 16.5 hours of observations and videotapes recorded at NWOC took 5 to 10 hours per hour of observation. Videotapes are first viewed and segmented into units for encoding. The size of the unit depends on the goal of the analysis. At times, a single sentence can be divided into several segments for coding. Conclusions drawn from the extracted data are encoded separately. The NWOC tapes resulted in 9057 encoded segments. A few sections of the tapes could not be encoded due to technical difficulties; these were compared with 2397 segments of USN data that were re-encoded. An attempt was made to match the USN data with the RAN data for the expertise level. Both groups included one highly experienced forecaster.

The coding scheme and level of detail depend on the hypotheses being investigated. Coding schemes are developed by an iterative process to ensure reliability. Two or more knowledgeable encoders independently analyze the same tape. These analyses are then compared using a Delphi procedure whereby the encoders discuss disagreements in their independent codings and agree on the correct encoding for each disputed instance. This leads to a refined definition or a modification of the encoding scheme. Once the coding scheme is stable, i.e., inter-rater-reliability for independent encodings reaches an acceptable level (usually greater than 75% agreement) and there are no categories that are consistently coding differently, the remainder of the protocols are encoded. Coding schemes can, and usually do, include more than one level or variable. For example, the METOC data were encoded for cognitive task stage, data type, and data source/type. (These levels are defined in the section 6.) Analysis of the encoded protocols is focused on examining significant patterns. Pattern indicator categories include fixed sequences, relative duration and frequency, and cross-level correlations. Statistical methods can then be applied to these patterns.

6. RESULTS

The stated goal of this study was to map the information usage of the METOC decision maker/forecaster to information visualization tools, and to discriminate the effects that are dictated by the forecaster's tools and training versus those that are due to basic human cognition. This was performed by comparing the mappings of USN and RAN forecasters. For each level of analysis (workflow, CTA, and tool/task), this section reviews the USN data, reports the RAN data, and compares these data with the USN data. The results reported here are taken from the nine videotaped protocols totaling 16.5 hours.

6.1 WORKFLOW ANALYSIS RESULTS

Workflow analysis of the USN data indicates that there are three task threads that weave through the workflow (see figure 5(a)). While the order of these threads is not fixed, there is a primary sequential pattern. The goal of this process depends on current missions and conditions. For example, in the experimental scenarios conducted at NPMOC and shipboard, the mission

was to provide a forecast to support strike planning. Initially, the forecasters concentrated on the weather thread, which began with an understanding of the current conditions and then moved to forecasting future conditions. During this period, an iterative process occurred between the observational data and the forecast models. The forecasters then used their understanding of the current and predicted weather to employ a set of TDAs to assess the effects of weather conditions on the performance of sensors and weapons. Lastly, they created a brief to present the forecast to the users, decision makers, pilots, etc. Each thread used specific information and outputted specific products.

Workflow analysis and conversations with RAN forecasters indicated that the same basic threads were followed. Of the three threads, the weather thread is of special interest to this investigation because it is the one that employs the forecaster's expertise to provide added value above and beyond the computerized output from sensors, observations, and models. In no instance did the forecasters simply accept a single source or model for making their predictions.

Differences between the RAN and USN workflow are largely a function of the current mission. For example, as no tactical operations were underway, the TDA thread was not followed during the observation period. Therefore, the TDA thread in figure 5(b) that depicts the RAN workflow has been shaded.

Both the inputs and outputs are affected by the mission, available tools, and the end-user's needs. Thus, while USN forecasters provided target, tanking, en route, recovery, and sensor predictions weather for a strike mission, RAN forecasters provided local observations for the national database and local TAFs, as well as weather forecasts for projected operational areas. As the RAN forecasters were providing routine rather than mission-specific METEOC services and products, their activities were more scheduled and the threads more clearly separate.

The most striking difference between the USN and RAN workflow is that RAN forecasters produce paper weather charts similar to that in figure 6. In the U.S., charting is done electronically. Hand charting is more practical in Australia, where there are far fewer observation stations and large empty areas. Hand charting allows the experienced forecaster to evaluate every observation station. Where the data points are sparse, every observation is weighted heavily, and instrumentation or recording errors would distort the chart. Where data points are dense, which occurs in the U.S., the effect of a single faulty observation can be overcome by averaging or integrating with neighboring observation stations. (Note, however, that the averaging process can mask real variation and microclimates.)

16

(a) USN METOC Forecasters

Inputs

Three Task Threads

Outputs

Mission

Specific
  Warfare
  Needs

WWW
  Satellite,
  Models,
  Observations,
  etc.

RADAR

UAV

"Weather"

Model

⇕

Data

TDA

Presentation

Target Weather

Tanking Weather

En-Route Weather

EM/EO Predictions

BINGO Weather

(b) RAN METOC Forecasters

Inputs

Three Task Threads

Outputs

Specific
  Needs

WWW
  Satellite,
  Models,
  Observations,
  etc.

RADAR

Paper
  Charts
  Observations

"Weather"

Model

⇕

Data

TDA

Presentation

Observations

Current Conditions
  Chart

Forecasts
  Local, En-Route,
    Destination
    Conditions
  TAFs
  Special Requests

*Figure 5. Workflow Analysis of USN METOC Forecasters
and RAN METOC Forecasters*
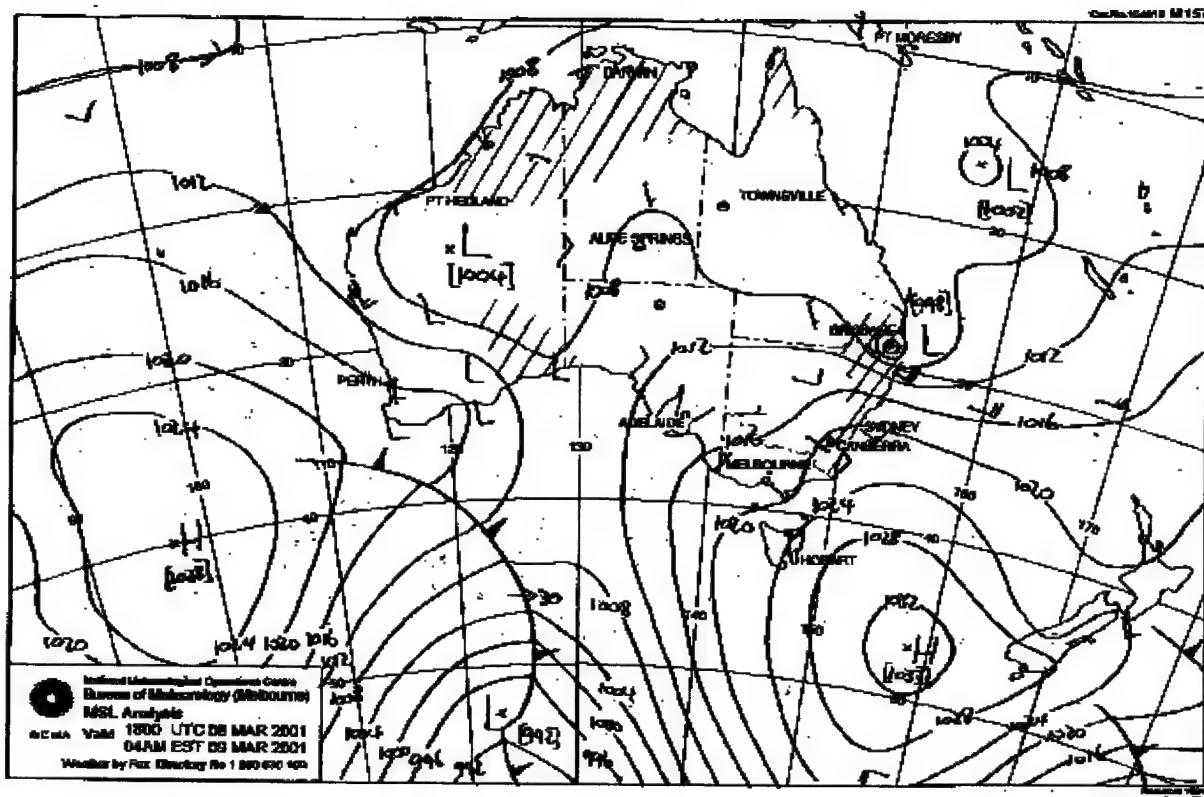
Figure 6. Example of a Hand-Drawn Weather Chart,
Such As Those Drawn by RAN Forecasters
(from the Australian Bureau of Meteorology web page, 9 March 2001)

6.2 COGNITIVE TASK ANALYSIS RESULTS

6.2.1 Cognitive Stages

Analysis of the USN data leads to a postulation of a four-stage cognitive model for the weather forecasting process.[11] These stages are as follows:

1. Initialize (update) general mental model. Initially, forecasters look for gross features, such as fronts, that will provide a context for the detailed forecast.

2. Build qualitative mental model (QMM). The forecaster develops an understanding of current conditions and how conditions are developing over the forecast period, and examines both current observations and models.

3. Verify and adjust QMM. At this stage, the forecaster compares his/her current understanding with that of other forecasters, models, satellite pictures, TAFs, etc.

18

4. Brief writing. During the brief-writing stage, the forecaster records and formats her/his forecast. Many of the forecast's parameters are reported with quantitative values.

5. Stages 1, 2, and 3 are a breakdown of the cognitive processes that are employed by the forecaster during the weather workflow thread. Stage 4 corresponds to the reporting thread in the workflow analysis. These stages are not a linear progression of activity. Forecasters iterate among them, especially between building and verifying the QMM. They will also make notes for the brief as they develop their forecast. Because these stages do not specifically address the TDA workflow thread, additional evidence for them can be found in the RAN data.

6.2.2   Information Usage

Analyzing information usage can provide two types of data. By showing similarities in classes of information usage that are independent of the tools, training, and teamwork patterns, support for the basic processes of human cognition can be found. In contrast, differences in information usage patterns, due to the impact of differences in tools, training and teamwork, can be imputed.

Two coding schemes that capture the way forecasters use information were developed to account for phenomena observed in these data. The first was developed for data collected in 1999 and 2000 at NPMOC.[11] This coding scheme is summarized in a somewhat simplified form in table 2. For each CTA stage of the forecasting process, it encoded the usage, the type of data being used (qualitative or quantitative), and data source. Usage encodings included setting a goal, extracting information, and brief writing (recording information to be used in the brief). Within the extract and brief writing categories, the type of data and data sources were also encoded. Data were classified as either quantitative or qualitative. The data sources included visualizations such as satellite pictures, charts, graphs, and text, and non-visualizations such as the forecaster's internal QMM and knowledge schema. These were further categorized according to whether they were being integrated with other sources or used alone. Note that any of these encodings can be employed at any stage in the cognitive process. For clarity and simplicity, the three stages of the weather thread are combined in the table.

Two codings within the USN data, interrupt and instrumentation, are not reported here. Both are idiosyncratic to the settings where they were observed and thus should not be compared. Interrupts, in the USN setting, were primarily due to questions from the technician. Many of these questions related to locating data or working with TDAs, and were, more or less, requests for instruction. Therefore, the forecaster's responses were largely instructional in nature. Instrumentation events were often artificialities induced by the experimental setting and the added machine burden.

Table 2. Coding Scheme Developed for USN METOC Data

Usage	Type of Data	Source	Example
Goal Statement			Need to look at surface pressure and winds at the 500-millibar height.
Extract	Quantitative	Chart (JMV)	The low is at latitude 33.5 N and longitude 120 W.
	Qualitative	Chart (JMV) and Satellite Picture (integrated cognitively)	The low is right on the money on JMV compared to the satellite image.
	Quantitative	Graph (Skew-T)	The numbers are not here; will have to interpolate mentally at 15°.
	Qualitative	Chart (JMV) and QMM (integrated cognitively)	It (JMV) shows a bit more funneling there (compared to his expectations).
Brief Writing	Quantitative	Chart (JMV)	Wind speed is 20 knots.
	Qualitative	Satellite Picture	Skies are mostly cloudy.
	Quantitative	QMM	En route weather, 30,000-foot level, winds will be about 25 knots.

The second coding scheme is a re-evaluation of the first, and is slightly altered for the Australian data (table 3). A subset of the USN data was re-analyzed using these encodings. In the revised encoding scheme, the Extract coding was split in order to separate extracting information from a visualization, comparing information from two or more visualizations (Compare), and deriving information by combining what was available in the visualization with the forecaster's knowledge (Derive). Thus, Compare subsumes all cases of integrated visualizations. Derive is used only in cases where the expertise of the forecaster plays a role and where the resulting verbalization goes beyond what is explicitly available in the visualizations. Lastly, Search was encoded separately and the brief-writing label was changed to Record to differentiate between the workflow thread of creating a brief and the cognitive task. The distinction is a pragmatic one, motivated by differences in the rhythm of the work at the two sites.

**Table 3. Coding Scheme with Examples Derived from USN Transcripts**

Usage	Definition	Example
Extract	To read information from any visible source.	Looks like PVA over the area.
Compare	To use two or more sources and compare them to any data.	Radar shows precipitation, but I can't really see anything on the satellite picture.
Search	Looking for a data source.	Okay, looking for weather maps over the area from Monterrey.
Record	Recording information for reporting to users. It need not be the final form.	This is a good picture right here, I'll take this. Just crop this picture a little bit.
Derive	To combine visible information with knowledge, so as to come to a conclusion that is different from that which is in the visible source.	So NOGAPS (model) doesn't seem to be handling the system very well.

For the RAN forecasters, recording includes both observation information and forecasts. As with the earlier encoding scheme, visualizations were classified as either qualitative or quantitative in data type and encoded source(s). However, only Compare and Derive integrate more than one data source, which are either external visualizations or internal mental representations, respectively. Two other encodings, Interrupt and Misc, were used but not included in table 3 because they do not account for information. As noted above, due to the experimental settings, interrupts are not comparable to the USN data. Interruptions in the RAN centers were of two types: on task and off task. A phone call requesting weather information or the need to make a scheduled observation was encoded as an on-task interruption. A phone call or visitor discussing a non-forecasting topic was encoded as an off-task interruption. In one test of rating reliability (using a rating/re-rating method with greater than 30 days between ratings), these encodings were found to be reliable according to Cohen's Kappa (K corrects for chance matches), where $K = 0.53882$ and $z = 7.10843$.

6.2.3 Qualitative and Quantitative Information Usage

The encodings were analyzed for overall information usage and individual cognitive tasks. Of special interest was the finding that there is a reversal in data usage (quantitative versus qualitative) from the weather thread stages to the brief-writing stage. This finding, reported in Trafton et al.,[11] was replicated with the RAN data. As seen in figure 7, both the USN and RAN forecasters primarily used qualitative information while developing their mental models of the weather and forecasts. The Analyze and Building the QMM terms focused on different aspects

of the process but are synonymous; the same is true for the Brief Writing and Record Forecast terms. However, when developing their briefs, the forecasters recorded mostly quantitative information, and in many cases these numbers did not exactly match any of the models that they had consulted. The numbers were apparently derived from the models, but modified by interpolation, the latest local observations, radar pictures, and experience. Replication of this reversal is a clear indication of the qualitative nature of the mental model. The effect also confirms the conjecture that forecasters rely on integrating a significant amount of complex information, not just the recollection of specific parameter values.

(a) USN Forecasters   (b) RAN Forecasters
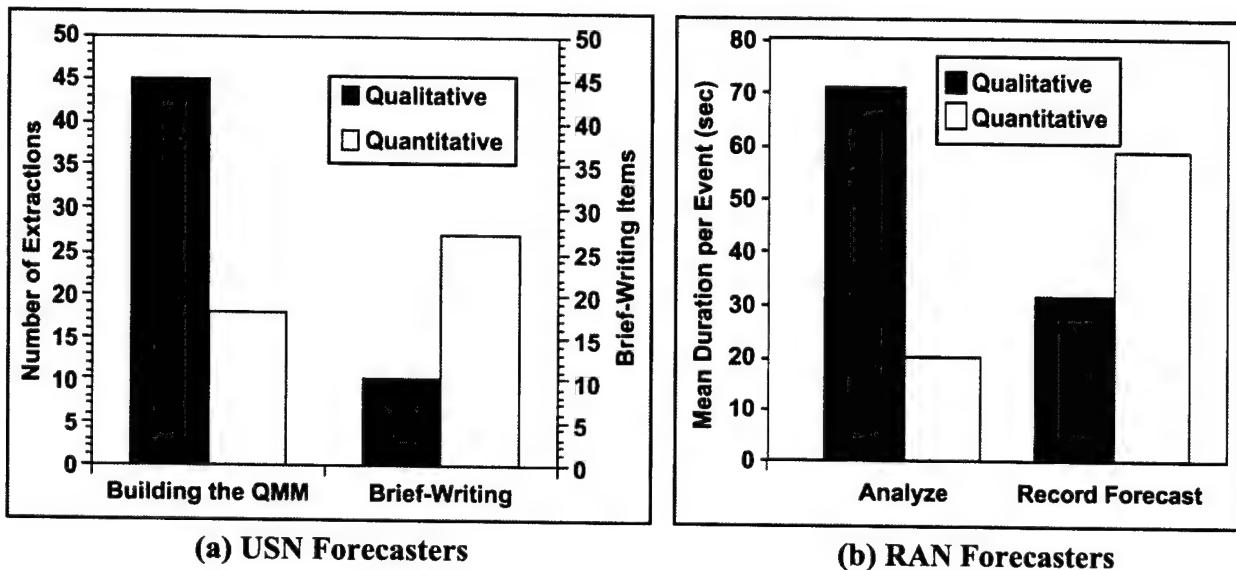
Figure 7. Use of Qualitative and Quantitative Information During the Weather and Recording Threads for (a) USN Forecasters and (b) RAN Forecasters

6.2.4 Information Retrieval and Usage Processes

While figure 7 indicates a strong similarity between USN and RAN qualitative and quantitative information usage, figure 8 shows the differences in the details of how these forecasters perform their tasks using the resources at hand and within their own specific environments (weather, training, and manning). The basic processes are the same, and there were no methods that were used by one group and not by the other. However, the relative frequency with which these methods were used show significant differences in some areas.

Figure 8 is a detailed breakdown of how the forecasters do their job: extracting information from visualizations, comparing visualizations, deriving information, and recording the forecast. Because each observed session took a different amount of time, the results are recorded as relative frequency, i.e., the proportion of total activity. The relative frequency of extracting, comparing, deriving, and recording information is of great interest. First, the USN forecasters appear to spend a larger proportion of their time extracting information. This difference is not statistically significant due to large variances. In contrast, RAN forecasters

22

spend a significantly larger proportion of their time comparing information (Mann-Whitney $U = 10.000$, $p = 0.053$, and $\chi^2(1) = 3.750$). They also appear to spend a slightly, but non-significantly, larger proportion of their time deriving information. The tools that RAN forecasters use support comparisons because there are adjacent monitors. Thus, they can see a model and satellite or radar picture, Skew-T, or other information visualizations. They can also examine two models side by side on the same monitor. USN forecasters must extract information, store it in memory or on paper, and then make comparisons from memory. From the results reported in the previous section, the assumption can be made that memory storage is usually in the form of a qualitative mental model and not a literal recall. Thus, the only exact storage is the paper notes made in conjunction with the extractions.
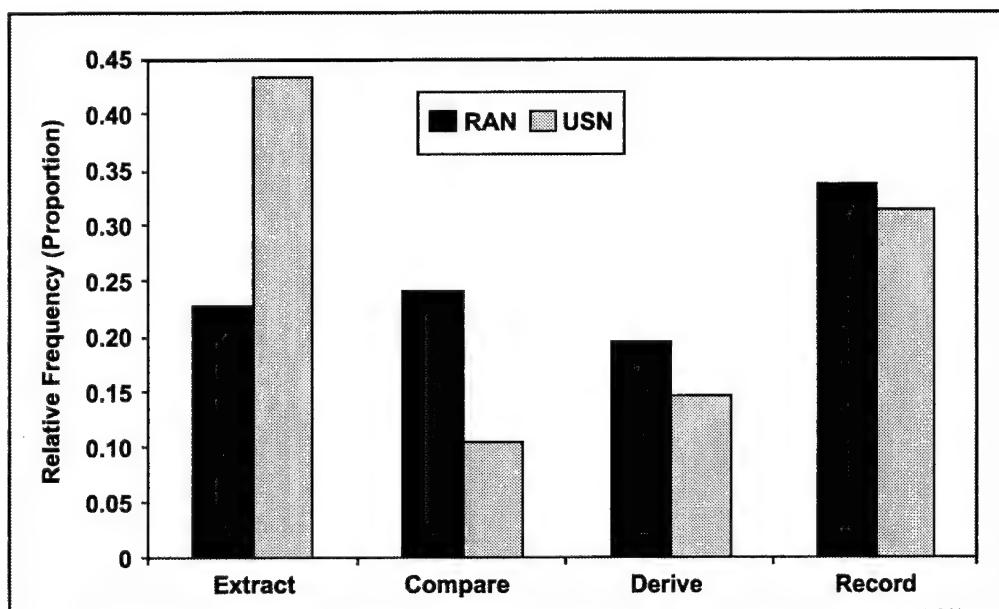
Figure 8. Relative Frequency of Cognitive Data Processes for All Forecasters

There are no differences in the data for the relative frequency of Record actions. However, there is one difference that was evident from the video data: the USN forecasters made two kinds of records for briefing purposes, one from unstructured paper notes as memory aids and the other from computer records using PowerPoint (the paper notes included data that were subsequently used as inputs for TDAs or for inputs into the PowerPoint brief), while the RAN forecasters created redundant sets of records for both observations and forecasts. One set was recorded on paper and one was recorded electronically for access by users (pilots, etc.). The paper version meets a legal requirement and is signed and archived. For briefs, the NWOC forecasters used the paper charts and pre-formatted plastic sheets for briefing. The computerized forecasts went into TAFs and a local system called the Centralized Airfield Coordination and Common Task User System (CACCTUS). The FWOC forecasters brief from paper weather charts but also use PowerPoint, when available. (A PowerPoint brief could not be examined due to technical difficulties with the software.)

6.2.5 Expertise Effects

The RAN and USN journeymen forecaster's backgrounds and training are very different. As noted above, RAN forecasters are officers with a college degree and at least 1 year's professional METOC training; USN forecasters are enlisted with a minimum of one METOC course. This course is at a lower level in part because the students enter with lower qualifications. The U.S. Navy emphasizes on-the-job training rather than a combination of academic and skill training requird by the Australian BoM course. As the journeymen are so different, it is worth comparing the USN and RAN experts. Both experts have college degrees, METOC certification, and extensive experience. Thus, they are quite comparable in both training and experience, although working in different environments and with different toolsets.

Figure 9 shows the relative frequency of cognitive tasks for the RAN and USN experts. As these are individual events, it is impossible to test for statistically significant differences and difficult to determine the reasons for the apparent differences. The differences may be due to the weather on the day in question or due to individual differences. Experts are more idiosyncratic in their performance than non-experts,[17] perhaps because they know more ways to perform the same task. Thus, it is instructive to examine these two as representative examples of how experts process weather data. Given the possibility for extreme differences, it is noteworthy that the methods used by the two experts are so similar. One possible reason why the USN expert performed relatively more extractions of information and the RAN expert spent relatively more time deriving information is that the weather during the USN observation was worse than during the Australian observation. Thus, the USN expert examined more models. The slight difference in the proportion of comparisons is probably due to chance.
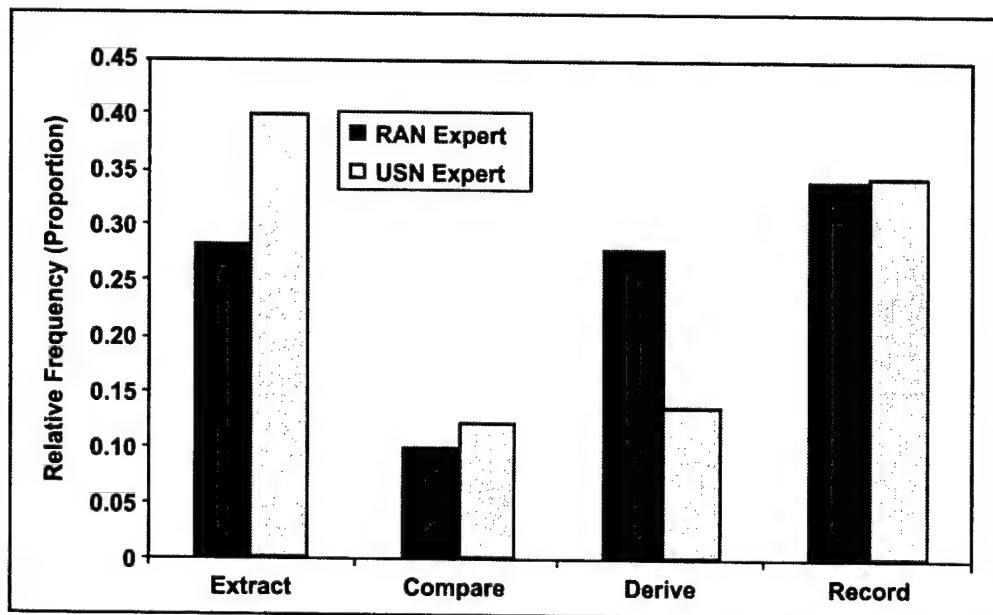

Figure 9. Relative Frequency of Cognitive Tasks for RAN and USN Experts

Another interesting way to examine these data is the probability of transitioning from one data task to another. Table 4 and figure 10 show that probability. For example, given that the RAN forecaster is currently extracting data, the probability of his next action being Comparing, Deriving, or Recording are $p = 0.11, 0.44$, and 0.44, respectively, which are represented by line weights in the figure. After eliminating all zero transitions, the remaining probabilities were ordered by rank. Line weights represent thirds of that ranking. These transition diagrams emphasize the importance of the Extract and Record poles. For both experts, more arrows both enter and leave those nodes of the diagram. Of the three node transitions, the most common cycles for both experts was either extract → record → extract, or record → extract → record. For the RAN expert, the extract → derive → record cycle was also common. Transitions between Compare and Derive are noticeably fewer than those involving the poles. This supports the conjecture that these two processes are complementary.

Lastly, these cognitive processes can be compared by stage (figure 11). The differences between Compare and Validate are isolated to the Validate stage. The USN expert can directly compare his forecast with the opinion of the METOC center by communicating over a chat channel. In contrast, the RAN expert must validate his forecast by comparing it with his own weather understanding; thus he Derives the forecast accuracy. Note that the RAN expert spends a considerable proportion of his time recording during every stage, which is probably due to the need to record everything twice, once on paper and once electronically.

Table 4. Probability of Going From/To Process
(Weight correlates with line weight in figure 10.)

From/To	Extract	Compare	Derive	Record
RAN				
Extract		0.11	**0.44**	**0.44**
Compare	0.40		0.20	0.40
Derive	0.31	0.08		**0.62**
Record	**0.48**	0.24	0.29	
USN				
Extract	0.08	0.25	**0.42**	0.25
Compare	0.33		0.33	0.33
Derive	**0.50**	0.13		0.38
Record	**0.60**		0.20	0.20
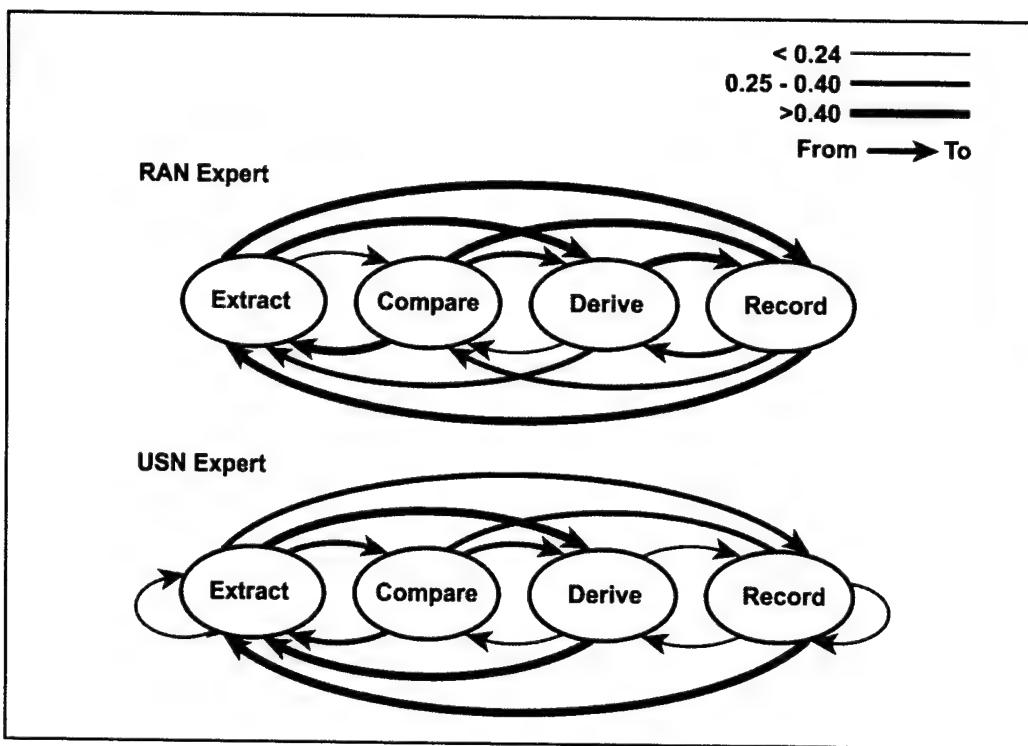
Figure 10. Probability of Transitioning from One Processing Task to Another for RAN and USN Experts
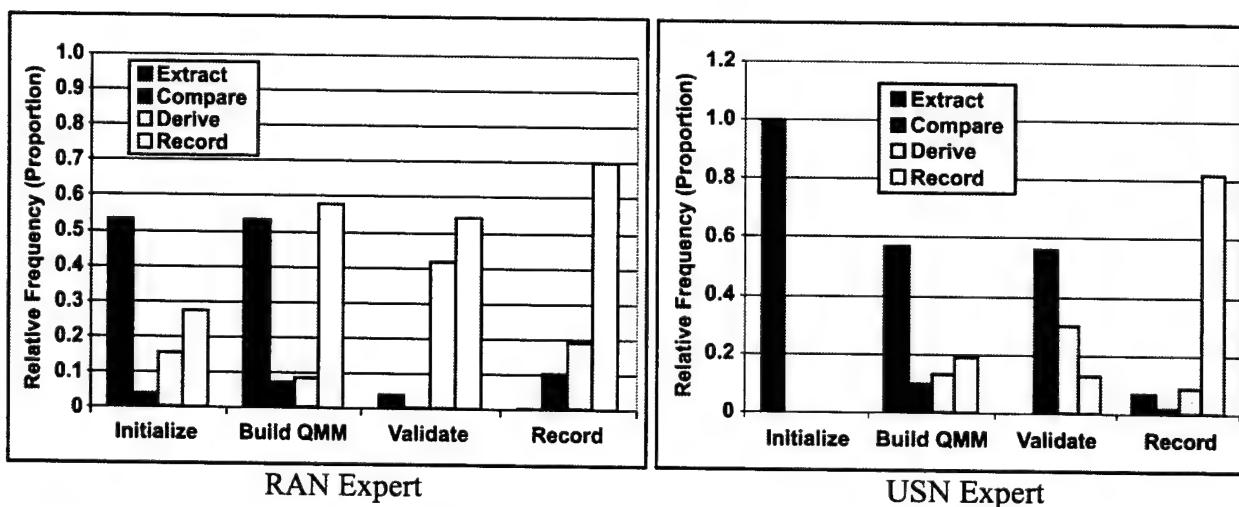
Figure 11. Relative Frequency of Each Cognitive Process by Stage for Expert Forecasters

26

As noted above, the transitions between Compare and Derive are noticeably fewer than those involving the poles. If a comparison is made between the RAN expert and non-expert data (figure 12), one striking difference is that the expert spent a greater proportion of time deriving information while the non-expert was required to compensate by making direct comparisons. This suggests that the expert is able to rely on memory to a greater extent than the non-expert. Findings from other domains such as chess[21, 22] support the conjecture that experts have a better memory than non-experts within their area of expertise. These results also suggest that support for comparison is especially important for journeymen forecasters. The role that comparisons play in the assessment of uncertainty is addressed in the following section. Note that the non-experts observed in this study were not novices; they had considerable training and experience and should be classified as journeymen forecasters. For example, they did not employ the same evaluation strategies as found by Lowe,[23] which entailed dividing the Australian continent into western and eastern parts. They all used the expert-like strategy of dividing their analysis into northern (tropical) and southern (temperate) sections.
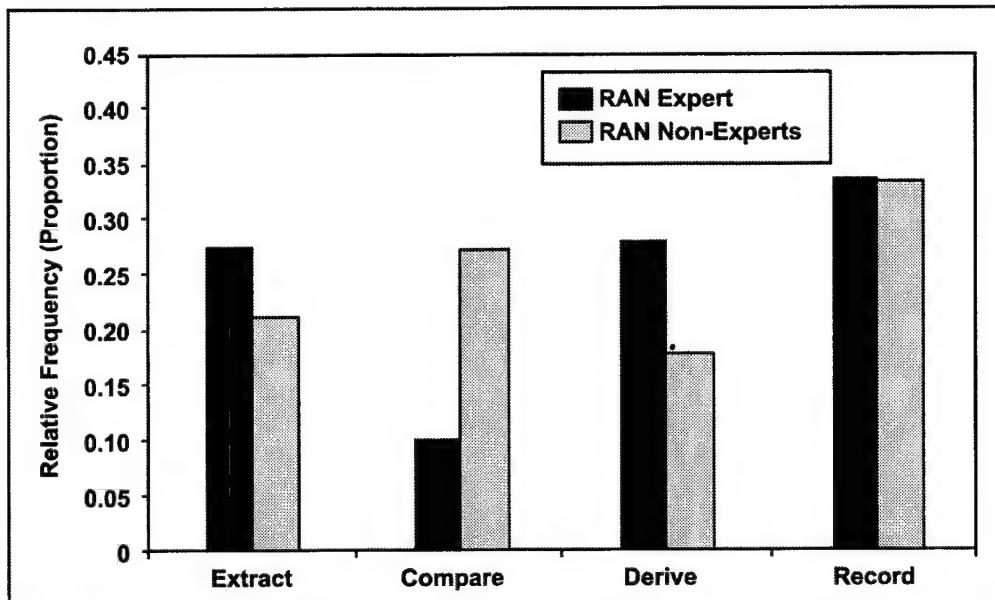
Figure 12. Relative Frequency of Cognitive Tasks by Expertise Level (RAN Only)

6.3 HUMAN FACTORS AND TOOL/TASK ANALYSIS

Tool/task analysis provides a metric for usability and is based on instances when the user must change his or her focus from doing the task to adjusting the tool. Thus, the tool/task ratio measures the proportion of total time spent manipulating the tool rather than using it to further the task. For the RAN forecasters, the observed mean = 0.15, which is considerably better than the mean of 0.26 observed in the USN data. The difference is probably due to the differences between the experimental and operational settings, and should not be taken as an indication of a USN norm.

27

The most prevalent tool-only actions observed were searches, which were a very small proportion of the total actions and were generally a search among the forecast models for the desired time. These searches had to be duplicated because they were often associated with comparisons. Thus, both models had to be adjusted to the same time. Although consuming little time, a mechanism that allowed the linked scrolling of two windows (containing the two models for comparison) would have accelerated the process and eliminated the occasional error that occurred when the two windows were not correctly synchronized.

One class of human-factors issues that the tool/task ratio does not address is the duplication of actions. For the RAN forecaster, the primary source of duplication is the requirement to record the same information both on paper and electronically. This requirement applies to both observations and forecasts. While the requirement to sign a paper copy of the observations or forecast certifies that the documents represents the individual's observation/forecast, a signed printout of the electronic document would fulfill the legal requirement. Furthermore, a printed rather than handwritten document retains legibility, even if the originator is no longer available. One indication of the savings that might be possible is the large proportion of Record actions, even while building the QMM. Approximately 33% of the Record actions were on paper, which accounted for more time than the computerized Record.

6.4 EXTENSION: UNCERTAINTY

As is well known, there are no deterministic models of weather today. All forecasts are derived by probabilistic models and are based on observed current conditions and assumptions about how these conditions relate to future conditions. Although these are probabilistic models, there are no explicit indications of uncertainty in any of the visualizations, forecasting tools, or data displays observed on the two continents. However, the differences in predictions are evidence of their inexactitude. To evaluate the models, the forecasters in both the USN and RAN make many individual comparisons between model predictions for the current conditions and the actual conditions (e.g., satellite loops) and among the models. These comparisons appear to help the forecasters assess the uncertainties in the models. The first type of comparison assesses the accuracy of the model at the current time. The second type assesses the amount of uncertainty in the prediction by determining the extent to which the models differ. Agreement among models increases confidence. Disagreement requires adjustment in the forecaster's predictions.

Comparisons are made on many variables, across time, altitude, and geographic location. Figure 13 is a photograph of a forecaster making a comparison between two models. In figure 14, the forecaster used a ruler to mark a fixed location as the prediction model looped through a time sequence. Other examples abound in the corpus of the data. They include comparisons of the path of thunderstorms from two radar locations and comparisons between: meso-scale and synoptic-scale models; models and satellite loops; models and balloon data; and many others. Overall, RAN forecasters spend approximately 25% of the total time on the task engaged in these comparisons. USN forecasters spend less of their total time on direct comparisons, perhaps because the tools provided by the experimental setting do not support side-by-side comparisons. When analysis of data from shipboard observations is complete, there may be evidence for a greater proportion of comparisons by USN forecasters as well.

28
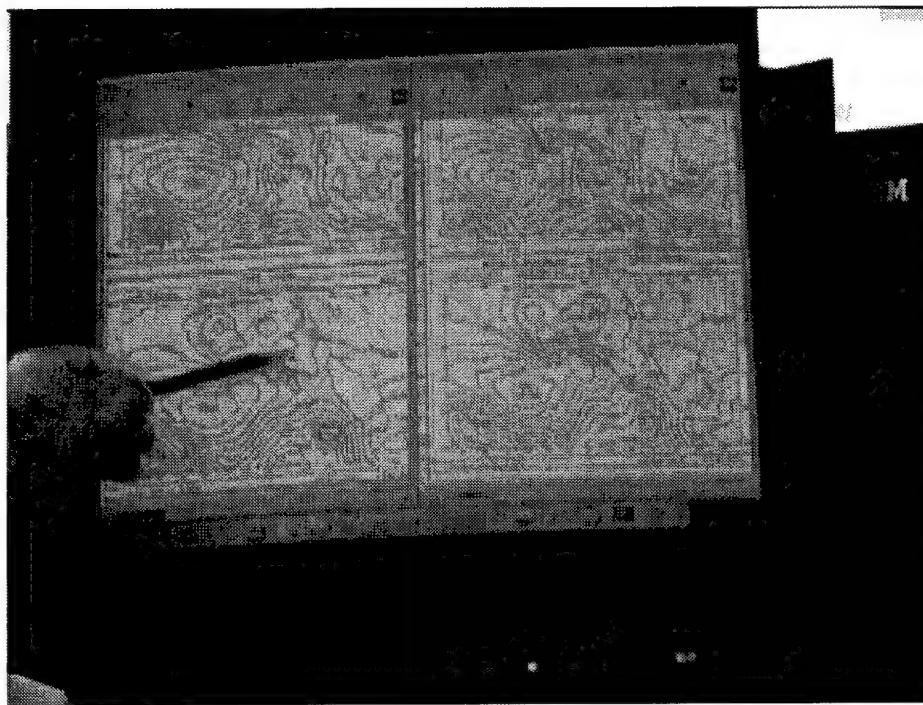
Figure 13. Forecaster Comparing the Same Feature As Predicted by Two Models
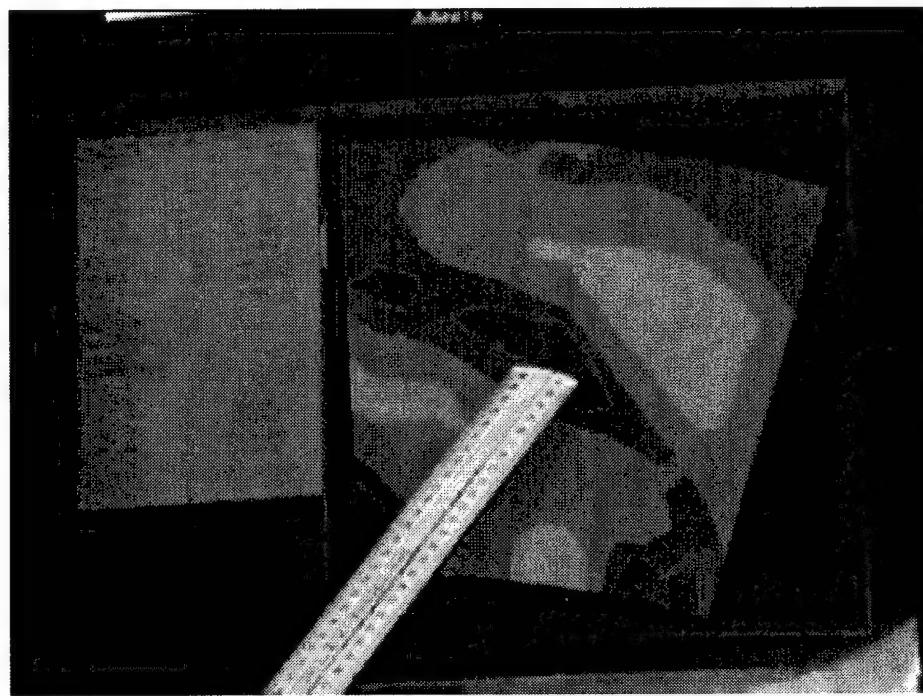
Figure 14. Forecaster Comparing the Same Location over Time
(Note the use of a marker to maintain the location across multiple frames.)

The comparison process of uncertainty evaluation and resolution supports the development of the QMM and the Validation process. Figure 15 shows the relative frequency of each of the information processing tasks during each of the four stages. Note that only during the Validate stage do Compare and Derive actions significantly outweigh other processes. However, even during report writing, there is considerable re-evaluation being performed as evidenced by the Compare and Derive actions. The presence of so many qualitative Compare and Derive actions during the brief-writing stage is in contrast to the more quantitative material being produced, which indicates uncertainty and continued reliability assessment.
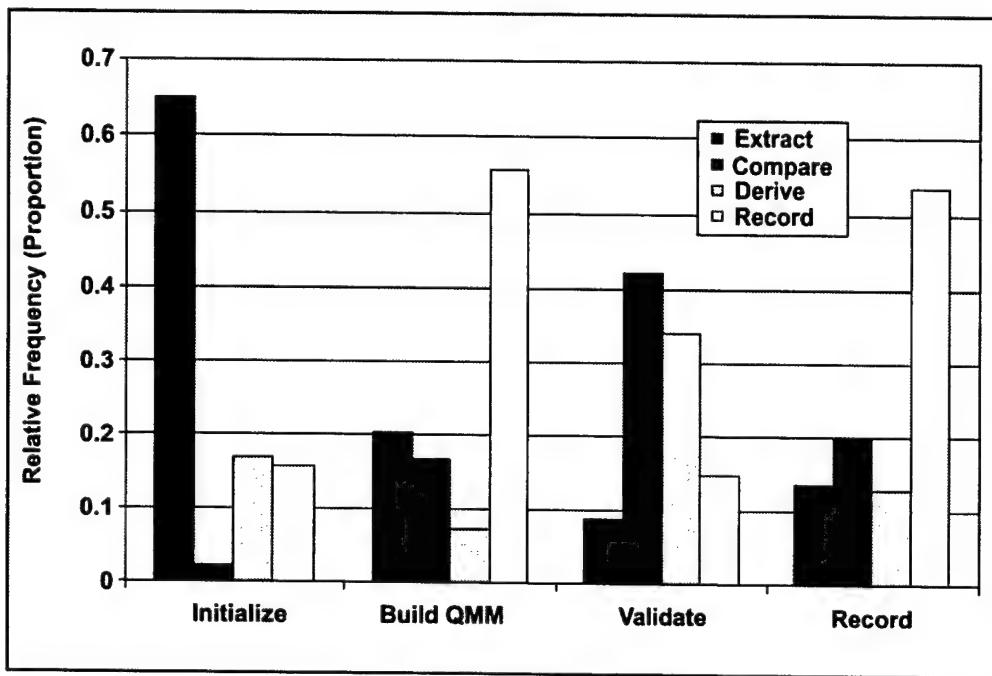
Figure 15. Cognitive Processes by Stage for All RAN Forecasters

6.5 DESCRIPTIVE PROCESS MODEL

By extending the work of Trafton et al.[11] with additional evidence from the RAN observations, the descriptive model of the weather analysis/forecasting/reporting process can be refined. First, an examination of the information processing procedures employed during each of the four stages (i.e., Initialize (update) general mental model, Build QMM, Validate (and adjust QMM), and Record (brief writing) must be performed. The tables below are taken from the protocols of individual forecasters and are typical of sequences observed in all nine protocols.

Table 5 is a typical example of the sequence of actions from one forecaster. Forecasters are always aware of the weather and make weather observations even when off duty. They also note the accuracy of their forecasts on an ongoing basis. The few Derives (17%) arise from this self-evaluation. Initialization varies slightly, depending on whether the forecaster is addressing a known and continuing area of interest or attending to a new region for which he or she has not recently forecasted. The forecaster quickly looks at only one or two representations (extract), compares them to his or her general mental model and/or derives where the attention should be directed, and may make a few notes (records). Even when addressing a new region, this takes only a few minutes. For example, the following text, taken from one transcript, is the complete initialization for that forecaster that day (table 5 is from another protocol): "Have a look at the pressure trends around the plains. Um, obviously rising pressure flow. Pretty much what drives the layers sometimes."

Table 5. Typical Sequence of Actions in the Initialize Process

extract ....extract.....→ derive → extract....extract....→ compare → record → exit

Initialization is fast (24 to 105 seconds), and information is dominated by Extract processes (65% of actions during the Initialize phase). Building the QMM accounts for the bulk of the forecaster's time (53%). It occurs with both current weather analysis and forecasting for the future. An extended model of this process must emphasize the iterative nature of the *extract → compare/derive → extract/record* cycle. Some of the variants of that cycle (taken from the protocol of one forecaster) are shown in table 6. Variants 1 and 2 are the most common. However, the others are also common variants. At this time, the control structure that selects among variants cannot be defined. Note that Building the QMM virtually always begins with extracting data from some visualization. As noted earlier, the majority of these extractions are qualitative in nature. Many of the visualizations do not support quantitative extractions (e.g., a satellite picture). In other instances, the forecasters discuss the data in qualitative, not quantitative terms. Many of the sequences end with recording, with either graphics (e.g., chart) or text. Variant 6 was observed during the end of the period while the forecaster was still Building the QMM. Subsequently, the forecaster went through a series of Validation and Recording cycles.

Table 6. Variants of Building the QMM

Variant	Sequence
1	*extract → derive → record*
2	*extract → record → extract*
3	*extract → derive → extract → record → compare → record*
4	*extract → compare → derive → record → extract*
5	*extract → compare → extract → derive → record → extract*
6	*record... record...→ compare → record .... record →* exit

Unlike the previous two stages, the Verify and Adjust QMM stages (table 7) do not generally begin with extracting new data, but are more likely to begin with comparing the current QMM to some other source. Often these are verbal discussions with either the incoming forecaster as shifts change or with the OIC. The two forecasters are likely to extract data together and derive a joint understanding from it. Adjustments to the forecast are then recorded. In contrast, the USN forecasters could not directly discuss their forecasts. However, the expert did use the Internet chat facility to discuss his forecast with the center.

Table 7. Typical Verify and Adjust QMM Sequences

1	*derive → derive → record → derive → extract → compare → record →* exit
2	*compare...compare...compare → record →* exit

Lastly, during the Brief-Writing stage, cognitive activity is heavily weighted by recording, although there are Extract, Compare, and Derive actions. A typical sequence (table 8) is dominated by Record actions. The sources of the occasional Extract, Compare, and Derive actions include the forecaster's QMM, notes made during other phases, and the paper record of the same data. Record sequences can be as long as 12 actions, and may include recording the same data, first on paper and then in the computer.

Table 8. Typical Recording Sequence

record....record → compare → record....record → derive → record....record → exit

In summary, this expanded understanding provides insight into the nature of the cognitive processes that support initiating, building, and validating the QMM, and then constructing the report (paper, online, or live brief). While these expansions to the model of the naval weather forecaster fill in some of the details, they do not completely specify the cognitive process. In particular, the expansions do not account for the control mechanism that determines when the forecaster moves among stages. What has been learned should allow better support for the progression from qualitative to quantitative, and from uncertainty toward confidence.

7. CONCLUSIONS AND RECOMMENDATIONS

Observations and data analysis have led to several recommendations. These include strengths of the RAN tools and training that could be emphasized and brought to the USN METOC community, in addition to improvements in the RAN METOC processes. While the author is not METOC qualified, she has spent 3 years observing forecasters and has 18 years experience analyzing the interaction among tasks, supporting tools, and the human factors (training, teaming, workload, and cognitive processes). The techniques of this analysis, workflow analysis, CTA, and tool/task analysis have been successfully applied to many domains. These recommendations will hopefully result in improved efficiency, reduced errors, and reduced workload. It is the author's hope that the recommendations provided here will provide similar results for both the USN and RAN METOC communities.

7.1 STRENGTHS

7.1.1 Organization and Operations

Although small, the RAN METOC organization has several strengths. The small size means that advancements in technology or knowledge can be quickly integrated into the centers. Conversely, problems that the forecasters encounter can be brought to the attention of support personnel, technologists, or management with a minimum of bureaucracy. The training, professionalism, and coordination with the Australian BoM compensate for the small size. However, the ability to fill all the billets would enhance the organization's ability to perform its mission, especially in time of inclement weather or operational stress. The lack of personnel is overcome by long hours and the excellent use of technology. By virtue of co-location with the principal customers and professional training, the RAN forecasters are able to tailor their forecasts to the needs of their customers. Because they are co-located and interact on a regular basis (as well as informally) with their customers, the forecasters are able to anticipate specific questions. Additionally, because the forecasters are certified and many have advanced training, they enjoy credibility with their customers.

At NWOC, the forecaster briefs each squadron separately. This might be seen as inefficient, as much of the material is repeated at each briefing; however, the daily contact with these small groups allows the forecaster to tailor the briefings to the needs of each squadron and its current operations. As the center is located in the tower building, it is convenient for individuals needing updated information to stop by at times other than the scheduled daily briefs.

At FWOC, fewer briefings are given because the number of interested groups is less than at NWOC. However, the briefs are no less tailored to the needs of the tactical decision makers. In order to accomplish this tailoring, the forecaster must know the current and planned schedules for all ships in order to provide the required information. Again, co-location and accessibility help maintain the necessary working knowledge of the situation.

7.1.2   Facilities

Both METOC centers have sufficient computing resources and supporting layout of the equipment. The lighted chart table and large monitors aid in visualizing the data and performing routine chores, such as comparing the current and previous charts or viewing visualizations. The two-headed computer with side-by-side monitors in use at NWOC facilitates comparisons. Large screens even allow the forecasters to view two models (two windows) simultaneously, thereby facilitating comparisons. Unfortunately, as the windows are not linked, keeping them synchronized requires considerable scrolling.

7.1.3   Hand Charting

While hand charting is time consuming, it is also a practical and effective way for the forecaster to build a detailed QMM of the current conditions. It is therefore the basis for building the future forecast. No forecaster objected to drawing it, and the more experienced forecasters were the most likely to express respect for the understanding that they gained from the act of hand charting. The charts also formed a strong background for forecasting and briefing. The down side to hand charting is that hand-drawn charts are not preserved or accessible to users outside the facility where they are created; scanning and digitizing them would alleviate this problem.

7.1.4   CACCTUS

The CACCTUS system is an effective way to communicate with the squadrons at NWOC. It allows the forecaster to update a previous forecast without retyping items that have remained the same. It sends the forecast to the same system that the squadrons use for other purposes. Thus, it is readily at hand. The system could be used in lieu of, rather than in addition to, handwritten forecasts.

34

7.2 WEAKNESSES

7.2.1   Handwritten Records

One of the weakest areas of the RAN system is the continued use of handwritten forecasts and observation records. This is a duplication of effort because the forecasters currently send the observation data and forecasts electronically to various users. Elimination of handwritten records would reduce the current tool/task ratio of 0.26 significantly. The handwritten records have a number of disadvantages:

Handwriting legibility varies widely. Once the forecaster who made the original record has left the facility, it can be very difficult for someone else to read the original. For example, the author copied both the forecasts and observations from the days when she observed, but found it impossible to interpret them later. (Thinking the difficulty was due to the use of acronyms and specialized vocabulary, the author sent copies, with the identification removed, to a recognized meteorological expert, but he was also unable to read them.) If signed copies are required for legal purposes, forecasts and observations could be printed from the electronic media and signed (electronic signatures are legally becoming more accepted).

Some disadvatages to handwritten records are as follows:

1. Paper records are costly to archive, and it is more difficult to retrieve an individual record when required. Electronic records can be searched by date, originator, or contents, and can be stored inexpensively and sent easily where needed.

2. Handwritten records that are transcribed into computerized form is an unnecessary duplication of effort for an organization that is already undermanned. Observations, in particular, are copied by hand from an electronic readout of sensor data.

3. In addition to being a duplication of effort, the copying and recopying by hand provides opportunity for error, either in copying or reading the handwritten record. There were several instanaces where the author personally observed the forecaster, while charting, noting erroneous observations that were mislocated, outside reasonable boundaries for the particular variable, or unreasonable, given the neighboring values.

7.2.2   Empty Billets

While the small size of an organization is not an overall weakness, the inability to fill all the billets is a problem that limits its effectiveness and places an excessive burden on the existing staff. Specifically, the lack of support at NWOC limits the forecasters' ability to collect observational data, not to mention the added burden placed on their ability to perform their tasks. When weather conditions required half-hourly observations, the forecaster's train of thought was interrupted frequently, regardless of the current task. This often required re-orientation after performing the observation, which further reduced efficiency.

7.3 ADDITIONAL RECOMMENDATIONS

The following recommendations are in addition to the need to remedy the weaknesses listed above. They are primarily ways to improve functionality and reduce the tool/task ratio and cognitive workload. These recommendations are not prioritized and no importance is implied by the order in which they are listed.

1. Support for comparisons. As comparisons are such an important part of the forecasting process, there are a number of enhancements to current visualizations that may be implemented. Currently, comparisons are the only method available to assess the reliability of models and observations; thus, they are critically important to forecasting. Two major recommendations in this area are as follows:

- When making comparisons between models, forecasters were observed to scroll excessively, from one window and to the other, in order to keep them synchronized. If a mechanism could be provided to synchronize the two windows, they could be scrolled jointly, thereby facilitating accurate comparisons.

- Further support for comparisons could be provided by enhancing the visualizations to emphasize the model's differences, for example, by differentially coloring the regions of disagreement. As the graphics are created from digital models, this enhancement is possible.

2. Continued developments. While there are many useful tools in both the current RAN toolset and under development, efforts should continue to take advantage of new technologies and ways to visualize data. The integration of weather and other types of data in a single visualization supports the RAN METOC mission. Most of these visualizations should continue to be geo-spatially and time (4-D) referenced for ease of communication with customers and interoperability with alliance partners.

Development of tools to visualize meso-scale models and the interplay of weather features with geographic features is especially important. While forecasters use what they know about the effects of local terrain (e.g., the ranges and valleys to the west of Nowra) on weather, they must also forecast for regions where they may not know the terrain. Because the METOC services cover any area where RAN ships and planes may be deployed, forecasters need to be able to visualize the interactions of weather and terrain in new areas as the need arises.
A second 4-D visualization that could be developed, in conjunction with operational groups, could show both tactical information and weather information. Thus, the location of ships (military and non-military) and tactical targets, in addition to stationary sites such as harbors and airfields, could be noted on weather/forecast maps. For the customer, this makes the impact of weather on operations an emergent feature.

36

The development of reliable information for models and observations can help forecasters be more efficient and accurate. Currently, models are assessed, but the results of the assessments are only provided to forecasters as general guidance. There are no indicators provided concurrently with the visualizations that show statistical variance, the age of the input observations, or the compliance of input data with the modeling assumptions. Note that, to avoid clutter, such additional information should be a selectable feature of the visualization (i.e., one that could be turned on or off). Indication of the presence of a large variance could be accomplished by color coding, giving the forecaster the option of investigating further. This recommendation is provided last, not because it is less important, but because it requires significant algorithmic and computational effort to provide current uncertainty information for each model run, and to determine the best way to communicate this information to the forecaster.

This work has replicated the major findings on workflow patterns and the use of qualitative and quantitative information by naval weather forecasters using complex visualizations,[11] and has provided additional details about the stages of that process. By extension, the process of using complex visualizations to develop a QMM of the systems and then constructing a quantitative report from that process may aid problem solvers in other complex domains in understanding high-dimensionality in dynamic systems. There were no indications that the available visualizations and toolsets failed to support this qualitative/quantitative dichotomy.

In contrast to the confirming results for qualitative and quantitative information used during the forecasting and reporting phases, the results show the differing influence of tools, training, or teamwork patterns on the more detailed aspects of forecasting. The clearest evidence addresses the question of how forecasters assess and compensate for the lack of precision and reliability in forecast models. Evidence comes from the extract/compare/derive tradeoff seen in the USN-RAN comparisons (figures 8, 9, and 11) and in the expert-novice comparison (figure 12). For example, overall the USN forecasters conducted more extract operations while RAN forecasters either compared or derived information. Both Compare and Derive operations included the extraction of relevant data, but extended the usefulness of the data by additional cognitive processing. This is an example of where the toolset supported deeper processing by facilitating integration of information; i.e., more data provides better information, not information overload.

Lastly, the results also show that experienced forecasters make extensive use of comparisons across models, between models and observations (including satellite and radar loops), and across time to assess the undisplayed uncertainty in observations and models. The detailed sequence model for the Verify and Adjust stages emphasizes how important the assessment of uncertainty is for the forecaster. This is an area that requires significant additional work, but one that could have a significant payoff in both the accuracy of forecasts and the amount of effort required to produce them.

Weather forecasting is a complex task requiring training, experience, and a specialized toolset for visualizing a multi-dimensional and interrelated data set. The findings reported here suggest that forecasters construct a qualitative mental model of the specific weather from quantitative and qualitative data, from their own knowledge of weather systems, and from how

the systems interact with the terrain. After validating this QMM to accommodate or reduce uncertainty, they could construct their largely quantitative forecast from it. This general descriptive process may well fit other complex cognitive tasks, such as submarining, astronomy, or reading medical images (for example, X-rays or MRIs). Additional research is already underway to investigate this problem. Regardless of the results, the study reported here has shown that this description of the cognitive process is accurate for forecasters who differ in training, tools, teamwork, and even the orientation of local weather systems, and has validated the model, at least for meteorologists.

8. REFERENCES

1. J. A. Ballas, D. W Jones, S. S. Kirschenbaum, J. G. Trafton, T. Tsui, and H. W. Wulfeck, "Human Centered Environmental Visualization and Tactical Weather Forecasting," *Proceedings of the Human Systems Integration Symposium*, Crystal City, VA, 4–5 November 2001.

2. K. A. Ericsson and H. A. Simon, *Protocol Analysis: Verbal Reports as Data*, The MIT Press, Cambridge, MA, 1984.

3. W. D. Gray, B. E. John, and M. E. Atwood, "Project Ernestine: Validating a GOMS Analysis and Predicting and Explaining Real-World Peformance," *Human-Computer Interaction*, vol. 8, no. 3, 1993, pp. 237–309.

4. P. M. Sanderson and C. Fisher, "Exploratory Sequential Data Analysis: Foundations," *Human-Computer Interaction*, vol. 9, nos. 3 and 4, 1994, pp. 251–317.

5. Y. Anzai and H. A. Simon, "The Theory of Learning by Doing," *Psychological Review*, vol. 86, no. 2, 1979, pp. 124–140.

6. H. J. M. Tabachneck-Schijf, A. M. Leonardo, and H. A. Simon, "CaMeRa: A Computational Model of Multiple Representations," *Cognitive Science*, vol. 21, no. 3, 1997, pp. 305–350.

7. E. M Altman and B. E. John, "Episodic Indexing: A Model of Memory and Attention Events," *Cognitive Science*, vol. 23, no. 2, 1999, pp. 117–156.

8. W. D. Gray, B. E. John, and M. E. Atwood, "The Précis of Project Ernestine or An Overview of a Validation of GOMS," *CHI'92 Conference on Human Factors in Computing Systems,* P. Bauersfeld, J. Bennett, and G. Lynch, eds., ACM Press, New York, 1992, pp. 307–312.

9. J. H. Larkin and H. A. Simon, "Why a Diagram Is (Sometimes) Worth Ten Thousand Words," *Cognitive Science*, vol. 11, no. 1, 1987, pp. 65–99.

10. E. Hutchins, *Cognition in the Wild*, The MIT Press, Cambridge, MA, 1995.

11. J. G. Trafton, S. S. Kirschenbaum, T. L. Tsui, R. T. Miyamoto, J. A. Ballas, and P. D. Raymond, "Turning Pictures into Numbers: Use of Complex Visualizations," *International Journal of Human Computer Studies, Special Issue on Empirical Evaluations of Information Visualizations,* vol. 53, no. 5, 2000, pp. 827–850.

12. W. Edwards, "Dynamic Decision Theory and Probabilistic Information Processing," *Human Factors*, vol. 4, 1967, pp. 59–73.

13. H. J. Einhorn, "Judging Probable Cause," *Psychological Bulletin*, vol. 99, no. 1, 1986, pp. 1–17.

14. D. Kahneman and A. Tversky, "Choices, Values, and Frames," *American Psychologist*, vol. 39, 1984, pp. 341–350.

15. A. S. Elstein, L. S. Shulman, and S. A. Sprafka, *Medical Problem Solving: An Analysis of Clinical Reasoning*, Harvard University Press, Cambridge, MA, 1978.

16. A. H. Bellenkes, C. D. Wickens, and A. F. Kramer, "Visual Scanning and Pilot Expertise: The Role of Attentional Flexibility and Mental Model Development," *Aviation, Space and Environmental Medicine*, vol. 68, no. 7, 1997, pp. 569–578.

17. S. S. Kirschenbaum, "Influence of Experience on Information-Gathering Strategies," *Journal of Applied Psychology*, vol. 77, no. 3, 1992, pp. 343–352.

18. P. M. Sanderson, J. J. P. Scott, T. Johnston, J. Mainzer, L. M. Watanabe, and J. M. James, "MacSHAPA and the Enterprise of Exploratory Sequential Data Analysis (ESDA)," *International Journal of Human-Computer Studies*, vol. 41, no. 5, 1994, pp. 633–681.

19. J. M. Schraagen, S. F. Chipman, and V. L. Shalin, eds., *Cognitive Task Analysis*, Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 2000.

20. W. D. Gray and S. S. Kirschenbaum, "Analyzing A Novel Expertise: The Unmarked Road," in *Cognitive Task Analysis*, S. F. Chipman, J. M. Schraagen, and V. L. Shalin, eds., Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 2000, pp. 275–290.

21. W. G. Chase and H. A. Simon, "Perception in Chess," *Cognitive Psychology*, vol. 4, no. 1, 1973, pp. 55–81.

22. K. A. Ericsson and J. Smith, eds., *Toward A General Theory of Expertise: Prospects and Limits*, Cambridge University Press, New York, 1991.

23. R. K. Lowe, Components of Expertise in the Perception and Interpretation of Meteorological Charts, in *Interpreting Remote Sensing Imagery: Human Factors,* R. R. Hoffman and A. B. Markman, eds., Lewis Publishers, Boca Raton, FL, 2001, pp. 185–206.

METOC Headquarters, Sydney

Initial visit was conducted on 18-19 January. I met all the key players and was introduced to the tools that forecasters, weather observers (including untrained sailors at sea), and data managers/validators use. These tools have been developed and are continuing to be developed in-house, in collaboration with industry or with the National Bureau of Meteorology (BOM). All development is with the guidance of the naval forecasters.

Fleet Weather and Oceanographic Center (FWOC)

22 Jan 2001: Observed most of day's 12-hour watch (except the preparation of a classified brief). Activities include preparation of a weather chart that is used for forecasting and preparation of forecasts for all ships for the period 2100 today to 2100 tomorrow. This amounts to making forecasts for next ~30 hours. The major time sink.

All observations were recorded in MacShapa. They can be reviewed and analyzed back at Swinburne next week.

23 Jan 2001: Observed early morning weather brief to staff and preparations of charts. The charts that are drawn are large scale, the size of the drawing chart. Charts are drawn every 6 hours. The forecaster was the same as yesterday. She was supervised by another forecaster with 2 years experience.

24 Jan 2001: Observed new forecaster while preparing his first brief. He used page-size sheet w/outline of land areas to draw two charts, the fairly current (000z) analysis and the prog for +24 hours after 000z. (Note: 000Z amounts to about +4–6 hours.) These hand-drawn sheets are scanned and turned into PowerPoint slides. The two slides and the satellite loop are the only visual aids that the forecaster uses when briefing. There are no text slides. There is additional concern about the weather for tomorrow as there will be a large outdoor inspection and ceremony (called "Division").

NAS Weather and Oceanography Centre (NWOC) HMAS Albatross, Nowra, New South Wales

Observed at NWOC for 1 week. Here forecasters support RAN air squadrons in both training and performing operations. HMAS Albatross is the only RAN NAS, and thus NWOC supports all air operations, including ASW and strike, as needed. NWOC is located in the building at the base of the tower and is accessible to air traffic controllers.

12 Feb 2001: Observed forecaster from about 0830. Forecaster has about 1 year's experience. He prepared chart and made 4-day forecast for the region. He also took hourly observations, recorded them, and transmitted them to other interested sites. He then turned the observations into an updated TAF. The system used is called CACTUS and was locally developed. It allows all squadrons (and other authorized users) to call up the forecast on their own computers. This system has cut phone requests for forecasts significantly. The primary interruptions are for phone calls. These include requests for specific forecasts, administrative issues, and other issues. At midday the shifts changed with a turnover brief. The second forecaster continued to make observations hourly and update the TAF. He also did a second chart and upper atmosphere analysis. There was a relatively informal afternoon brief at 1600 for squadrons that would be flying in the evening. Unlike the morning brief, in which the forecaster goes to each squadron in the afternoon, the squadron leaders, ATC, and others came to the weather center.

13 Feb 2001: Began observations at 0800. The forecaster this morning is less experienced than yesterday's forecaster. She completed the year-long course several months ago and has been forecasting at NWOC for about 3 months. When I arrived she was conducting squadron briefs. Tomorrow I will arrive in time to observe her preparing for the briefs (4:00 AM) and hope to accompany her on these. She had already done her morning chart, but I did observe her doing the four-day forecast. She specifically mentioned comparing models early on in the taping. I don't know if I got it on tape because she has a quiet voice. I added the lapel mike after that. Observed turnover and then participated in visit by USN visitors including John Kendal. We received a tour of the squadrons here at HMAS Albatross. These three helicopter squadrons and one jet squadron are the principal customers for NWOC forecasts. One of the helicopter squadrons fly Squirrels and is principally for training new pilots. The other two have Sea Hawks and Sea Kings. They have varied missions. The jet squadron is from New Zealand.

14 Feb 2001, 4:00 AM: This morning I arrived early enough to observe preparation for the morning briefs and the briefs themselves. This morning was interesting because the weather is changing and therefore required considerable more thought. There was early morning drizzle and the threat of more. I have a copy of the forecast at each of the stages. At the briefs, several of the squads asked about weekend weather in Melbourne because there is an air show there and several squads are flying to or from there. The forecaster knew about their interest and was prepared with the information that they wanted.

Recorded the handover brief and am staying for afternoon activity, including chart preparation, forecast, and afternoon brief. During the afternoon the workload got quite heavy because the cloud cover lowered and they went into special conditions. This requires half-hourly observations. The forecaster was also concerned about a line of thundershowers that were headed for or near the area. These storm cells were observed on radar and the airfield was on the edge of the system. What made prediction difficult was that the cells were observed while they were across the mountains from HMAS Albatross. Additionally, the timing of the system was difficult to determine due to a computer failure. The best radar system had not been working since the move to the new facility.

15 Feb 2001, 6:00 AM: Arrived in time for review of morning forecast with OIC prior to morning briefs. Also observed one brief and preparation of morning forecast. As the forecaster was very tired and unable to give a verbal protocol, I did not record the preparation of the mid-morning chart. Chart preparation was interrupted by the requirement to perform the hourly observation. A special request was received by telephone to forecast for tomorrow morning for both here and divert fields. This request was because of a flight that would be arriving with limited fuel. If conditions would be difficult, they would plan differently, either in terms of scheduling or fuel loading. Taped a discussion on predicted conditions for the target time. Taped handover. This afternoon I taped forecasting, observations, and weather discussions. During all taping, I turned the tape off for personal phone calls, conversations, and other irrelevant time.

16 Feb 2001, 9:38 AM: Observed early morning brief. Today is a really an odd weather day. The sky is very changeable, alternating between blue sky and clouds. This is a beautiful location for viewing both the sky and the airport activity. Today there will be an added flight of New Zealand planes coming to join the squad that is here.

A comment about the working environment at NWOC: At the end of my last day, shortly after I left for the airport, one of the NZ planes crashed in the woods, very near the airfield. There were no apparent adverse weather conditions, but I am sure that the single forecaster on duty was exceedingly busy documenting weather conditions at the time of the crash. This is an example of the kinds of exceptional workload (physical and emotional) that argue for additional staff. Having spent a week there, I was quite shaken by the news. I have come to think of the forecasters as friends. In both centers, the forecasters work alone. The only opportunity to validate the forecast is at shift changes or in discussion with the Officer in Charge. At NWOC, there are no support personnel either.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
Office of Naval Research (D. Jakubek (Code 44), Schmidt-Nielson (Code 342))	2
Naval Research Laboratory (J. Ballas, G. Trafton (Code 5532), E. Tsui (Code 7540))	3
Space and Naval Warfare Systems Center (N. Gizzi, W. Wulfeck)	2
Combat System Research Centre, Maritime Operations Division (T. Mansell)	1
RAN METOC Services, Maritime Headquarters (C. Roy)	1
METOC Futures, Performance and Capability (B. Dubsky)	1
Fleet Weather and Oceanography Centre (FWOC), Maritime Headquarters	1
Naval Air Station Weather and Oceanography Centre, HMAS Albatross (J. Couch)	1
University of Washington–APL/ETSD (D. Jones, R. Miyamoto)	2
San Diego State University–CERF (S. Marshall)	1
Defense Technical Information Center	2
Center for Naval Analyses	1